2018-11-16

Investigating the Contribution of Disordered Domains to the Biological Activity of RNA-binding Proteins

Davide Tavella
University of Massachusetts Medical School

INVESTIGATING THE CONTRIBUTION OF DISORDERED DOMAINS TO THE BIOLOGICAL ACTIVITY OF RNA-BINDING PROTEINS

A Dissertation Presented

by

DAVIDE TAVELLA

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

November 16^{th}, 2018

Biochemistry and Molecular Pharmacology

INVESTIGATING THE CONTRIBUTION OF DISORDERED DOMAINS TO THE BIOLOGICAL ACTIVITY OF RNA-BINDING PROTEINS

A Dissertation Presented
by
DAVIDE TAVELLA

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

---
Francesca Massi, PhD, Thesis Advisor

Celia Schiffer, PhD, Member of Committee

Osman Bilsel, PhD, Member of Committee

Brian Kelch, PhD, Member of Committee

Vincenzo Venditti, PhD, Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

---
Lawrence Stern, PhD, Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school

---
Mary Ellen Lane, PhD
Dean of the Graduate School of Biomedical Sciences

Program in Biochemistry and Molecular Pharmacology

November 16^{th}, 2018

DEDICATION

To my parents, Silvana and Roberto

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Dr. Francesca Massi, for having given me the independence to explore my scientific interests, while still guiding me with her useful advices. Also, I have been very lucky to work every day with awesome colleagues, thank you Laura and Asli. I really want to thank all the members of my thesis committee, Dr. Larry Stern, Dr. Celia Schiffer and Dr. Konstantin Zeldovich for tracking the progress of my research over the years and providing valuable suggestions. I sincerely thank Dr. Sean Ryder and Dr. Jill Zitzewitz for being amazing collaborators and for their suggestions during the past years in our weekly joint lab meetings. I also want to acknowledge everyone in the BMP department, it has really been a privilege to have worked with such talented people.

Special thanks go to my parents and to my brothers, for their unconditional love and support, and to my "Worcester Family": Raziel, Alex, Nicholas and Giuseppe. At last, I want to thank Giulia, my love and inspiration in life, science and everything else.

ABSTRACT

Many proteins contain disordered domains under physiological conditions. These disordered regions may be functional, although under pathological conditions they may lead to protein aggregation and degradation, as observed in proteins related to neurodegenerative diseases. In my thesis study, I aimed to understand how the primary sequence of these proteins encodes for the diverse ensemble of conformations rather than a stable folded state. I focused on the role of disordered domains in the activity of RNA-binding proteins involved in post-transcriptional regulation, but may lead to pathogenesis in many diseases.

The human TIS11 proteins bind to AU-rich elements in the 3′ UTR of mRNAs through a CCCH-type tandem zinc finger (TZF) domain. Mutations in these proteins have been linked to cancer. A member of this protein family, Tristetraprolin (TTP), is partially unfolded in the C-terminal zinc finger in the apo state, but folds upon RNA binding. The homolog protein TIS11d is folded in both free and bound states. Previous studies have shown that the extent of structure of the TZF domain in the apo state does not affect the affinity to target RNA in vitro, however it modulates the activity of the protein in cell. To understand which interactions determine the zinc affinity of the C-terminal zinc fingers of TTP and TIS11d, I investigated the stability of their TZF domains using homology modeling and molecular dynamics (MD) simulations. I found that, in the C-terminal zinc finger of TIS11d, a hydrogen bond is necessary to allow for π-π stacking between the side chains of a conserved phenylalanine and the zinc-coordinating histidine. Using mutagenesis and nuclear magnetic resonance (NMR) spectroscopy, I demonstrated that the lack of this hydrogen bond is responsible for the reduced zinc affinity, and thus lack of structure, of the C-terminal

zinc finger in TTP.

These results suggest that the CCCH-type TZF domain in different proteins have evolved to differentiate their function through a disorder-to-order transition.

In *Caenorhabditis elegans* several RNA-binding proteins contain a TZF domain homologous to the RNA-binding domain of TIS11 proteins, but have different RNA-binding specificity. I characterized the structure and the dynamics of the *C. elegans* protein MEX-5 using NMR spectroscopy and MD simulations. I found that MEX-5, like its mammalian counterpart TTP, contains a zinc finger that is partially unfolded in the free state but that folds upon RNA-binding. To assess if the disorder-to-order transition upon RNA-binding contributes to MEX-5 function, I designed a variant MEX-5 where both zinc fingers are stably folded in the absence of RNA. I characterized the RNA-binding activity of this variant MEX-5 and I found that the binding affinity and specificity are unchanged compared to the wild type protein. Together with Ryder's lab, we used CRISPR-hr to introduce this variant into the endogenous *C. elegans mex-5* locus. Homozygotes animals are sterile, form massive uterine tumors within a few days of reaching adulthood, and often die by bursting.

These results show that the unfolded state of MEX-5 is critical to its function in vivo by a mechanism distinct from its RNA-binding activity.

To further investigate how the equilibrium between structural order and disorder affects the function of a protein in the cell, I focused on the human protein TDP-43, a major component of the cellular proteinaceous aggregates found in amyotrophic lateral sclerosis and other neurodegenerative diseases. Previous studies have shown, both in vitro and in vivo, that the second RNA recognition motif (RRM2) of TDP-43 domain contains peptide regions that are particularly prone to fibril formation. In addition, RRM2 has been shown to populate, to a small degree, one or more partially folded states under native conditions. To determine if the partially folded states of TDP-43 RRM2 contribute to the formation of aggregates observed in the human dis-

eases, I characterized the structures of these states using MD simulations including enhanced sampling methods and restraints from experimental chemical shifts. I found that in these states the protein exposes to the solvent aggregation-prone regions that are instead buried in the protein core in the native state.

These results suggest a role in fibrogenesis for the transient partially folded states of TDP-43 RRM2.

PREFACE

Chapter II has previously been published as:

Tavella, Davide, Laura M. Deveau, Troy W. Whitfield, and Francesca Massi. "Structural basis of the disorder in the tandem zinc finger domain of the RNA-binding protein Tristetraprolin." Journal of chemical theory and computation 12, no. 10 (2016): 4717-4725.

Author Contributions:
DT, LMD and FM designed the experiments and simulations. DT carried out the MD simulations, analysis of the data and prepared and performed all the NMR and CD experiments of the T202L TIS11d mutant. LMD prepared the samples and performed all the NMR and CD experiments of TTP, wild type, F200A and C192S TIS11d. DT and FM wrote the paper.

Chapter 3 has been a collaborative study that is currently under preparation for publication as:

Tavella, Davide, Antkowiak, Katianna, Bulut, Reyyan , Ryder, Sean and Massi, Francesca. "Assessing the biological role of the disorder-to-order transition in the RNA-binding activity of the *C. elegans* protein MEX-5."

Author Contributions:
D.T., F.M. and S.R. designed the research. D.T. performed the simulations and

analysis, and prepared and performed all the NMR experiments. K.A. and S.R. performed the CRISPR/Cas9 genome editing, entablished the *C. elegans* mutant strain, and characterized the phenotypes. R.B and S.R. performed the microscopy imaging experiments.

Chapter 4 has previously been published as:

Tavella, Davide, Jill A. Zitzewitz, and Francesca Massi. "Characterization of TDP-43 RRM2 Partially Folded States and Their Significance to ALS Pathogenesis." Biophysical journal (2018).

Author Contributions:
D.T. and F.M. designed the research. D.T. performed the simulations and analysis. All authors wrote the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Disorder in protein and its function: intrinsically disordered proteins and partially folded states

The central dogma of structural biology has been, for roughly a century, that protein function associates only with a well-defined three-dimensional structure. This protein-structure-function paradigm was first introduced with the lock-and-key proposal by Fischer [1] and then continued with the interpretation of denaturation as loss of specific structure by Wu [2] and, independently, by Mirsky and Pauling [3]. Nowadays, it is accepted that proteins fall instead onto a structural continuum: starting from tightly folded single domains, to multidomain proteins that might have flexible or disordered regions, to compact but disordered molten globules and, finally, to highly extended, heterogeneous unstructured states [4, 5]. Proteins with intrinsically disordered regions (IDRs) or domains (IDPs) lack a sufficient hydrophobic core and therefore fail at folding spontaneously into a highly organized three-dimensional structures. However, IDPs and IDRs can adopt compact but disordered molten-globule-like states, or local regions of their sequence can have a propensity to adopt isolated and transient elements of secondary structure [4]. Consistently, it has been shown that proteins rarely behave as true random coils in aqueous solution and show propensity to form local elements of secondary structure or hydrophobic clusters even under denaturing condition [6].

The occurrence of unstructured regions is notably common in functional proteins. The existence of functional unstructured proteins have been longly acknowledged [7]

and unstructured proteins have been reported in intact cells since early proton nuclear magnetic resonance (NMR) spectroscopy experiments [8]. Compared to proteins that adopt a well-defined energetically stable three-dimensional structure, the free-energy landscapes of IDPs are relatively flat, allowing sampling of many different conformations. Reported examples of domains that directly use their structural flexibility to carry out their function are linkers [9], entropic springs [10] or elastomers [11], entropic bristles [12], and native molten globules [13, 14]. IDPs and IDRs are present in all organisms, but are particularly abundant in eukaryotes, where they play crucial roles in processes like signaling and regulation [15].

1.1.1 Functions of the intrinsic disorder in proteins

IDPs and IDRs facilitate several biological processes through their lack of structural constraints [5]. The features of IDPs allow a precise level of control of cellular processes. IDPs are characterized by the presence of small recognition elements that fold on partner binding and by a high degree of flexibility, which enables them to interact promiscuously with different targets in different contexts. Additionally, IDPs contain accessible sites for post-translational modification [16]. The ability of IDPs to bind partners with high specificity but modest affinity leads to rapid and spontaneous dissociation and termination of the signal [17, 15, 18] and kinetic advantages in signalling [19], as their extremely fast association rates allow signals to be rapidly turned on. The molecular interactions that occur in complexes involving IDPs are transient and dynamic. IDPs, in fact, exchange binding partners and compete for binding to central hub proteins, which are often present in limiting amounts [16]. These interactions are fine-tuned by post-translational modifications that enable them to function as switches and rheostats [15, 20, 21, 22, 23, 24]. Moreover, disorder-to-order transitions can underlie function [25], such as folding upon binding [26] or activation of chaperones following unfolding [27]. Many of these processes involve interactions of

IDPs and IDRs with partner proteins or nucleic acids.

The association between IDPs, or IDRs, and folded globular proteins can occur in the disordered form, through preformed structural elements or, alternatively, through conformational selection, first described for binding to a flexible ligand [28, 29]. Coupled binding and folding, or induced fit [30], represents a different mechanism for these protein-protein interactions, with structure formation occurring together with binding [31]. DNA- and RNA-binding proteins employ different strategies to overcome the thermodynamic and kinetic challenges of binding to specific nucleotide sequences: in order to form a network of interactions (i.e. hydrogen bonds) with the ligand molecule, the protein undergoes a conformational change that has an entropic cost. Several of these strategies include the interaction with unfolded or partially folded protein regions. Indeed, induced protein folding contributes to sequence-specific DNA binding with a large changes in heat-capacity, resulting from the conformational transition undergone upon the nucleic acid-protein complex formation [31]. Structural reorganization of the protein upon binding involves folding of entire domains, as in the basic region of the basic leucine-zipper DNA-binding domain [32], as well as folding of local disordered loops or linkers between folded domains, as in the Cys_2His_2 zinc finger proteins [33]. Additionally, several complexes involve a mutual induce-fit mechanism where both the protein and the nucleic acid are more structured in the bound state, as in the case of the ribosomal protein L5 associating with the 5S ribosomal RNA [34].

Furthermore, the structural instability and flexibility of the IDPs and IDRs might be exploited by the cell to target misfolded proteins for proteolytic degradation. Irreparably misfolded proteins are tagged for destruction in the cell by the ubiquitin-proteasome system. Targeting in the ubiquitin-proteasome pathway is driven by the identification of solvated hydrophobic amino acids. It has been recently shown that the yeast nuclear E3 ubiquitin ligase San1 uses disordered N- and C-terminal domains

to directly recognize misfolded substrates [35]. These IDRs contain short, conserved stretches of hydrophobic residues flanked by very flexible regions, which enable San1 to adaptively recognize many differently shaped misfolded substrates [35]. Although IDPs and IDRs exhibit features of unfolded proteins, many of them are resistant to proteolysis. The particular amino-acid composition, displaying linear sequence distributions of oppositely charged or polar residues (Gln, Ser, Pro, Glu, Lys) and the lack of bulky hydrophobic residues (Val, Leu, Ile, Met, Phe, Trp and Tyr) [36], may contribute to the requirement of intrinsically unfolded segments for long-term stability. Polyglutamine repeats, for example, have been found to be resistant to degradation by eukaryotic proteasomes [37].

1.1.2 Case studies: early examples, CBP

Many IDPs and IDRs with biological functions have been reported for more than 60 years, including casein [38], phosvitin [39], fibrinogen [40] and calcineurin [41]. Casein has been one of the earliest studied protein to exhibit peculiar features of unfolded protein: casein has been known, in fact, to tolerate prolonged heating and treatment with chemical denaturating agents without any significant structural change. Furthermore, casein is highly susceptible to protease degradation. In light of these observation, McMeekin in 1952 suggested that casein occurs in milk in an unfolded configuration, which may be rapidly digested by proteolytic enzymes. [38]. Recently, it has been proposed that casein molecules associate under physiological condition to form soluble micelles based on an IDP-based matrix [42]. These micelles are sustained by diverse types of interactions, including interactions of phosphorylated residues with calcium phosphate nanoclusters. Phosvitin, from egg yolk, was reported by Jirgensons in 1958 as a flexible polyanion [39]. The primary sequence of phosvitin, in fact, contains $\sim 57\%$ serine, mostly phosphorylated, many polar residues and very few hydrophobic moieties [43]. Fibrin is the cleavage product of the fibrinogen hex-

amer, a blood plasma glycoprotein, by thrombin protease. Once cleaved, fibrin forms an insoluble matrix [40]. Recent circular dichroism and NMR studies suggest that the region at the carboxyl end of the α-chain of fibrin contains structured regions of limited size that are connected by flexible unstructured linkers [44]. Calcineurin is a serine/threonine phosphatase activated by binding to calcium-calmodulin [45, 46]. In the crystal structures of calcineurin [47], the calcium-calmodulin binding site is located in a region of missing electron density. Moreover, hydrogen exchange experiments show that the binding site is unprotected in absence of calcium-calmodulin, supporting the hypothesis of its intrinsically disordered nature [48].

Intrinsically disordered regions are well exemplified by CBP. The transcriptional co-activator CBP, and its paralogue p300, have an intrinsic acetyltransferase activity that is used in turn to modify chromatin and transcription factors. Additionally, CBP functions as scaffolds for the recruitment and assembly of the transcriptional machinery [15]. Roughly 50% of the protein is intrinsically disordered, with folded domains connected by long unstructured regions containing predominantly Gln, Pro, Ser, Gly, Thr and Asn residues. One of the folded domain of CBP, the transcriptional-adaptor zinc-finger-1 (TAZ1) domain, mediates interactions with the hypoxia-inducible factor-1α (HIF1α), therefore regulating the hypoxic response [49]. The C-terminal transcription-activation domain of HIF1α is unstructured in solution, but undergoes local folding transitions to form three short helices on binding to the TAZ1 domain of CBP/p300 [15]. The tight binding ($K_d \sim 7$ nM) between the two domains represents an example of enthalpy-entropy compensation. There is a conformational entropic cost associated with the disorder-to-order transition that accompanies the binding of an intrinsically unstructured protein to its target. The key thermodynamic driving force for the binding reaction is usually a favorable enthalpic contribution, which gives an example of enthalpy-entropy compensation. Indeed, the coupling of folding and binding, in the case of HIF1α and CBP/p300, favors the burial

of an extremely large surface area, despite the relatively small size of the interacting domains [15]. The CBP/HIF1α system provides, also, several examples of how IDPs regulate their binding partners through covalent modifications. The residue Asn 803 of HIF1α, located in the disordered C-terminal transcription-activation domain, acts as a hypoxic switch: Asn 803 is hydroxylated in normoxic condition and disrupts the binding of HIF1α to CBP [50]. However, the enzyme responsible for Asn 803 hydroxylation requires that the surroundings residues assume an α-helix conformation, with the side chain of the asparagine projecting into the enzyme active site [51, 52]. The ability of the C-terminal transcription-activation domain of HIF1α to adapt its conformation to different binding partners, at a marginal energetic cost, arise from the intrinsic plasticity of the disordered domain. Finally, analysis of the CBP sequence offers insights on characteristic features of flexible linkers connecting folded globular domains. The amino acids composition of CBP linkers reveals a high proportion of polar residues, with a percentage, $\sim 70\%$, that is conserved in many species from *Caenorhabditis elegans* to *Homo sapiens* [15]. Interestingly, segments within the linkers of CBP contain stretches with relatively high proportions of hydrophobic and charged residues. These motifs, although not yet characterized, may represents binding sites for partner proteins [15].

1.1.3 Coupled folding and binding

Coupled folding and binding is the process in which an IDP, or IDR, undergoes a conformational transition from disorder to an ordered structure upon binding to its target. In this process there is an entropic cost to fold a disordered protein or region, which is paid for using the binding enthalpy. As proposed by Schulz in 1979 in its study on protein-nucleic acid interactions, this entropy/enthalpy compensation results in high specificity and low affinity [53]. However, recent studies have shown that, although many IDPs exhibit weak affinities with their partners, many do not and the

affinities for complexes involving IDPs span values common to complexes involving structured proteins [54]. Additionally, complexes involving IDPs are characterized by fast binding kinetics [55, 56], which may occur via a fly-casting mechanism [57] or a dock-and-coalesce mechanism [58]. In the fly-casting model, the unfolded domain binds weakly at a relatively large distance, followed by folding as the protein approaches the binding site. In the dock-and-coalesce model, intead, a segment of the IDP first docks to its cognate sub-site on the target, allowing the remaining segments to explore conformational space and coalesce around their cognate sub-sites. These features of IDPs complexes, high specificity and fast kinetics, play an important role in signal-transduction, in which proteins must properly associate to initiate the signaling process, but must be able to rapidly dissociate when signaling is completed.

Moreover, a system that includes disordered domains that fold on binding can take advantage of its conformational flexibility to facilitates the post-translational modification of the complex components. As discussed above for the CBP/HIF1α system, conformational flexibility confers to a protein the ability to interact with its physiological target and with modifying enzymes [15].

1.1.4 Methods and experimental techniques to study structural ensembles and dynamics

A central problem in studying IDPs and IDRs is to develop experimental techniques that allow to describe, at atomic resolution, free and complexed states of IDPs and IDRs. Traditional structural biology methods have been developed to characterize stably folded conformations of proteins. Therefore, these techniques are not appropriate to capture the structures and dynamics of IDPs and IDRs. Solution state NMR and molecular dynamics (MD) simulations, however, have proven to be suited to describe conformational ensemble of IDPs and IDRs and their structural disorder [59].

1.1.5 Nuclear Magnetic Resonance spectroscopy

NMR appears as a powerful tool to describe the conformational ensemble of IDPs and IDRs due to the specific averaging properties of diverse conformationally dependent observables [60]. NMR observables, such as chemical shifts, scalar couplings and residual dipolar couplings, in fact, are measured as population weighted average over all conformers if the system is in fast exchange regime, that is if structural interconversion happens in a timescale faster than the chemical shifts difference between the corresponding crosspeaks, usually milliseconds [61, 62]. In particular, chemical shifts report the chemical and electronic environment of the nucleus of interest. Deviations of measured chemical shifts from random coil values [63] provide important insights into the secondary structures populated in the conformational ensemble for IDPs or IDRs.

NMR relaxation allows to explore protein motions in terms of backbone and side chain dynamics. In the specific case of IDPs, NMR relaxation experiments provide description of local dynamic timescales, long-range reconfiguration timescales and conformational exchange (folding/binding) [62]. Standard techniques can be applied to partially folded proteins and IDPs, such as measurement of T_1, T_2 and heteronuclear NOE parameters for backbone resonances or R_2 relaxation dispersion [64]. Challenges in data analysis and interpretations, however, arise from the complexity of the theoretical framework for quantitative analysis of nuclear spin relaxation in terms of local and segmental motions [62].

1.1.6 Computational Biology (molecular dynamics and enhanced sampling)

Atomistic MD simulations represent an established technique for studying protein folding and the underlying free-energy landscapes. Some caveats in using MD simulations in explicit solvent for characterization of the structural ensemble of IDPs are the

computational power and the computational time required for adequate sampling of the conformational space. The use of enhanced sampling methods allow to overcome this issue. Enhanced sampling simulations are classified as unbiased (temperature replica exchange [65], solute tempering [66]) or biased (umbrella sampling [67], metadynamics [68]) [59].

In the temperature replica exchange MD method [65], multiple copies (replicas) of the system are simulated in parallel, all at different temperatures. While one replica is collected at the temperature of interest, the others are performed at a temperature high enough to rapidly overcome free-energy barriers between metastable states [59]. Exchanges between replicas adjacent in temperature are attempted during the simulations according to a Monte Carlo scheme [69]. Therefore, the conformational ensemble at the temperature of interest results from the fast sampling of conformational space of the high-temperature replicas.

In order to reduce the computational cost of MD simulations that effectively explore the conformational space, enhanced sampling methods that employ biasing potentials can be used rather than replica exchange-based approaches. In particular, metadynamics simulations exploit a biasing potential to discourage the system of interest to visit previously explored regions of the conformational space ([68]. During these simulations, Gaussian-shaped potentials are periodically deposited along the trajectory of a collective variable (CV), an appropriately chosen variable that describes the reaction coordinates. As a result, at the equilibrium the history-dependent biasing potential exactly compensate the unbiased free-energy profile along the chosen CV [59]. The advantage of metadynamics, compared to other biased methods [67, 70] consists in no required *a priori* knowledge of the end states. Additionally, multiple CVs can be used to describe the system and multi-dimensional free-energy landscape can be obtained [71].

The ensembles of conformations resulting from enhanced sampling methods are

analyzed by clustering the data based on structural similarities or calculating the probability of observables [59].

In this thesis, I have studied how the balance between structure and disorder in a protein affects its function. In particular, to address this fundamental question I have focused on three independent RNA-binding proteins: TTP, MEX-5 and TDP-43. In the following sections, I will describe the three proteins and their functions in the cell.

1.2 The role of CCCH-type Tandem Zinc Finger proteins in post-transcriptional gene regulation

The majority of the zinc finger proteins are DNA-binding proteins, specifically most of them are transcription factors. Among the family of zinc finger proteins, however, several members, such as the CCCH-type zinc finger proteins, bind and recognize RNA targets and are involved in regulation of the RNA metabolism, including mRNA splicing, polyadenylation, export, translation, decay, transcriptional repression or signaling modulation [72]. Structurally, CCCH-type zinc finger proteins are characterized by one or more finger motifs, each of them coordinating a zinc ion through three cysteines and a histidine.

Genome-wide surveys have revealed nearly 60 CCCH-type proteins in humans and mice [73], and similar results (~ 70) have been reported from genome analysis on *Arabidopsis thaliana* and rice [74]. These results suggest that CCCH-type zinc finger proteins are evolutionarily conserved across different species. Although the functions of most of these proteins have not been elucidated yet, to date CCCH-type zinc finger proteins have been reported to be involved in a broad range of biological processes, spanning from immune response (cytokine production, immune cell activation, immune homeostasis and antiviral innate immunity) to regulation of cell differentiation and cancer cell growth [75].

1.2.1 Gene regulation and ARE-binding proteins

The control of mRNA stability is an important process in the response of the cell to changes in intracellular and extracellular stimuli. This response requires rapid regulation of gene expression and mRNA translation, processes that are controlled by sequence elements that govern mRNA half-life, stimulating or inhibiting its degradation [76]. In mammalian cells AU-rich elements (AREs) are 50-150 nucleotides *cis*-acting sequences, within the 3′-untranslated regions (3′-UTR) of mRNAs, that

ARE classification	
Class I	Dispersed AUUUA in U-rich context
Class II	Overlapping $(AUUUA)_n$
Class II	U-rich stretch

Table 1.1: ARE Classes

have been reported to target mRNAs to rapid degradation [77]. AREs are defined not only by an enrichment in adenosine and uridine but, importantly, by their patterns, repeats, and context [78] (see table 1.1). Therefore, the minimum ARE motif is not trivially the AUUUA sequence, but rather is the pentamer in a AU-rich context (i.e. UAUUUAU or UUAUUUAU or overlapping repeats of the pentamer) [78]. The *cis*-acting sequences are specifically recognized and regulated by *trans*-acting factors, which can be RNA-binding proteins (RBPs) or noncoding RNAs, such as the microRNAs [78]. Specifically, to date several ARE-binding proteins have been studied and their mechanisms elucidated, such as the members of the CCCH TIS11 family, the AU-rich-binding factor 1 (AUF1) and the stability factor HuR [78].

Genes encoding ARE-containing mRNAs are usually involved in cell growth or response to micro-organisms, inflammatory stimuli and environmental factors [76]. These genes require very precise regulation of their spatial and temporal expression patterns, which in turn depend on transcriptional control, translation regulation and stability of the mRNA. In this context, AREs are essential elements to control the turnover of these mRNAs [79, 80, 81, 82]. In fact, under normal unstimulated condition, the ARE-dependent degradation mechanism represses very effectively the expression of these potent proteins. Dysregulation of ARE-mediated degradation has been reported to be associated with pathological states such as cancer, chronic inflammations and auto-immune pathologies [83, 84].

1.2.2 The TIS11 proteins family: Tristetraprolin and TIS11d

The TIS11 family of RNA-binding proteins includes three mammalian genes: Tristetraprolin (TTP), TIS11b and TIS11d [85, 86]. These proteins are induced in response to direct extracellular stimuli such as insulin, growth factors, mitogens and phorbol esters [85]. The TIS11 proteins bind to the class II AREs (see table 1.1) and consequently promote 5′-decapping or deadenylation and degradation of the mRNA via the recruitment of the exosome [87, 78] (see figure 1.1). TIS11 proteins have an important role in mediating the inflammatory response by regulating the stability of the mRNA of critical genes [85], such as TNF-α. Through recognition and binding to the TNF-α AREs, in fact, TTP, TIS11b and TIS11d control TNF-α concentration through the turnover of its mRNA [88, 89].

1.2.3 The CCCH-type Tandem Zinc Finger domain architecture

Proteins of the TIS11 family possess a RNA-binding activity by mean of a highly conserved tandem zinc finger (TZF) domain, containing two CCCH-type zinc finger elements [90]. The affinity of the zinc fingers to zinc is high, $K_d \leq 10^{-11}$ M [91], and mutations that abolish zinc-binding cause loss of RNA-binding activity [92, 93]. The primary sequences of the TZF domains in the TIS11 proteins highlight several common features (see table 1.2), including the conserved spacing between the ligand cysteines and histidine, with scheme $CX_8CX_5CX_3H$, and the length of the linker connecting the two zinc fingers, 18 residues. Other conserved elements of the primary sequence consist of the motifs preceding the first zinc-coordinating cysteine, (R/K)YKTEL (see table 1.2), that has been shown to be of pivotal importance to define the RNA-binding specificity [90].

In 2004, Hudson et al. solved the solution NMR structure of the TZF domain of TIS11d in complex with a cognate 5′-UUAUUUAUU-3′ RNA molecule [90]. The structure reveals a new fold for the CCCH-type TZF domain, with the two zinc fin-

TTP	SRYKTELCRTFSESGRCRYGAKCQFAHGLGELRQANRHPKYKTELCHKFYLQGRCPYGSRCHFIHNPSE
TIS11b	SRYKTELCRPFEENGACKYGDKCQFAHGIHELRSLTRHPKYKTELCRTFHTIGFCPYGPRCHFIHNAEE
TIS11d	TRYKTELCRPFEESGTCKYGEKCQFAHGFHELRSLTRHPKYKTELCRTFHTIGFCPYGPRCHFIHNADE

Table 1.2: Sequences of the TZF domains of TIS11 proteins. The TIS11 family proteins share homology at their TZF domain. The two zinc finger motifs (underlined) have a characteristic $CX_8CX_5CX_3H$ scheme of the zinc coordinating residues (highlighted in red). Conserved R(K)YKTEL sequences, preceding each finger, are highlighted in blue.

gers, showing the same backbone conformation (root mean square deviation = 0.84 Åfor the heavy atoms), connected by a the linker region. Each zinc finger adopts few secondary structure elements, specifically a short α-helix immediately after the first cysteine ligand and a turn of 3_{10}-helix between the second and third cysteine ligands [90]. Additionally, a turn of 3_{10}-helix is present at the interface between the N-terminal zinc finger and the linker [90].

1.2.4 The RNA-binding activity of Tristetraprolin and TIS11d

TTP and its paralogs TIS11b and TIS11d have been identified as RNA-binding proteins that lead to the destabilization of their polyadenylated mRNA targets containing ARE class II elements [94] via the recruitment of the exosome [87] or, alternatively, via the 5′ decapping and 5′ to 3′ degradation [87, 78] (see figure 1.1). Previous studies demonstrated that mutation of A to either G or C in the context of the ARE motif AUUUA abolishes, or strongly reduces, the ability of TIS11 proteins to bind such sequences [89, 95, 96].

The solution structure of TIS11d in complex with a cognate 5′-UUAUUUAUU-3′ RNA molecule provides useful insights into the binding footprint of these proteins on the ARE probe [90]. The structure of the complex, in fact, reveals that each zinc finger recognizes a single 5′-UAUU-3′ motif [90]. Interestingly, Hudson et al. also reported that for TIS11d, and likely for other TIS11 proteins, the ARE recognition is

accomplished through an extensive network of hydrogen bonds between the protein backbone and the bases of the RNA, mostly to the Watson-Crick edges [90]. The binding between protein and RNA is further stabilized by means of stacking interaction between the side chains of conserved aromatic amino acids and all of the RNA bases [90].

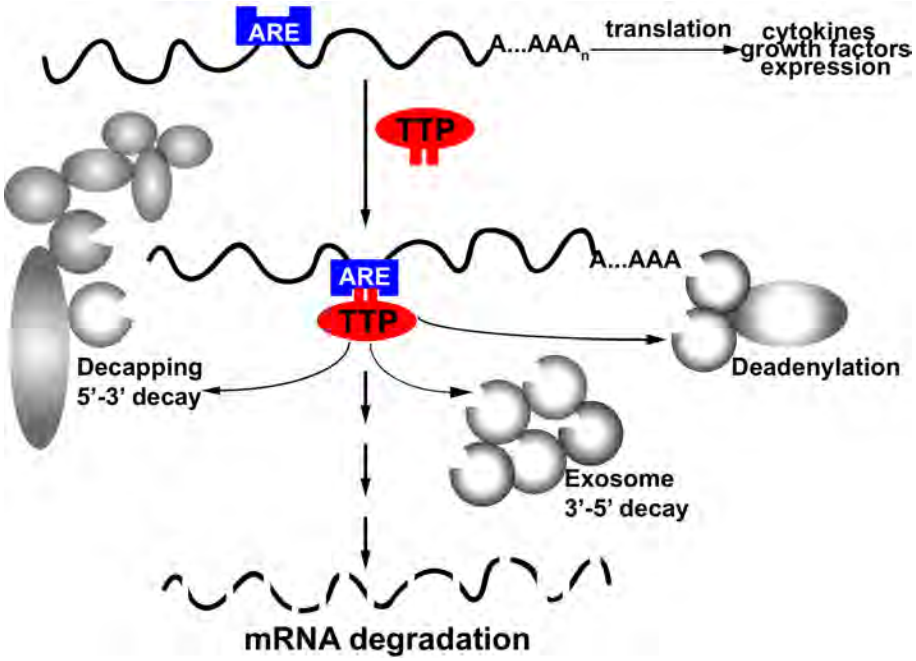


Figure 1.1: Post-transcriptional mRNA decay mechanisms. A model of mRNA degradation initiated by TTP binding to ARE-containing mRNA. Binding of TTP promotes degradation of the mRNA transcripts, thus decreasing protein production.

1.2.5 The role of TIS11 proteins in cancer

Members of the TIS11 protein family, by inhibiting translation of ARE-containig mRNAs, have a critical role in regulating cell growth, inflammation, angiogenesis, and invasion. Therefore, changes in expression levels, intracellular localization and activity of ARE-binding proteins can lead to cancer initiation and progression [97].

TTP regulates inflammation- and cancer-associated factors [97], and loss of TTP function is a common feature of several human malignancies, including brain, lung, ovary, prostate, thyroid, breast, cervix, colon and liver tumors (Human Protein Atlas available from www.proteinatlas.org).

TIS11d is a crucial modulator of hematopoiesis [98], and dysregulation of TIS11d function has been associated with leukemia pathogenesis [99]. Iwanaga et al. have identified several mutations of the *Tis11d* gene in patients with acute myeloid leukemia, acute lymphoblastic leukemia and in lymphoma cell lines [99]. Two of these mutations, P190L and D219E, occur in two highly conserved residues in the TZF domain of TIS11d. Interestingly, a previous work from our lab has shown that these two missense mutations cause differences in structural flexibility of the TZF domain of TIS11d and results in aberrant cytoplasmic localization [100]. These findings suggest a correlation between the loss of structure of TIS11d and its role in tumorigenesis.

Previous studies have shown a relationship between the activity in the cell of TIS11 proteins and the extent of structure of their TZF domains [96]. Moreover, two cancer-associated mutations in TIS11d have been demonstrated to cause loss of structure in its TZF domain [100]. To date, however, the molecular interactions that determine the stability of the CCCH-type zinc fingers are still not well characterized. The results presented in this thesis identify a stacking interaction between the zinc-coordinating histidine and a conserved aromatic side chain that is crucial for proper zinc coordination and stability of the finger fold.

1.3 The role of CCCH-type Tandem Zinc Finger proteins in early development of *Caenorhabditis elegans*

Caenorhabditis elegans is a widely used model system to investigate fundamental developmental processes of embryogenesis. One of the main theme of embryogenesis studies is the post-transcriptional regulation of maternal mRNAs, produced and reversibly silenced in the earlier stages of oogenesis. During the transition from oocyte to zygote, and for several cell divisions, the embryonic genome is transcriptionally quiescent [101]. Therefore, maternally encoded mRNAs and their translational regulation play a central role in the determination of cell fates and the establishment of embryonic axes [102].

1.3.1 Embryonic development in *Caenorhabditis elegans*

In the fertilized *C. elegans* oocyte, the point of sperm entry determines the posterior pole of the embryo, thus establishing the anterior-posterior body axis [103]. During the following stages of development, a series of rapid and invariant asymmetric cell divisions pattern the identity of the early embryonic blastomeres [104]. The first division of the zygote generates an anterior daughter cell (AB) and a posterior daughter cell (P_1). The AB daughter cell is the precursor of somatic tissues, whereas P_1 undergoes three subsequent asymmetric divisions, each time generating a somatic blastomere (EMS, C, and D) and a germline blastomere (P_2, P_3, and P_4) (see figure 1.2) [105]. Germline blastomeres are transcriptionally silent and progress more slowly than somatic sisters through the cell cycle. At the ∼100-cell stage, P_4 divides symmetrically to generate Z2 and Z3, which proliferate throughout larval development and give rise to the sperm and oocytes of the adult germline [105].

The initial patterning of the *C. elegans* embryo is controlled by maternal factors. Among these maternal factors, PAR proteins act as cell polarity regulators, orchestrating the anterior/posterior axis in the early *C. elegans* embryo [106, 107]. The 7

PAR proteins in *C. elegans* are categorized in two groups, based on their distributions in the anterior (PAR-3, PAR6 and the kinase PKC-3) or posterior (the kinase PAR-1 and the RING-finger protein PAR-2) cortical domains of the polarized zygote; two additional PAR proteins, PAR-4 and PAR-5, are symmetrically localized in the cytoplasm and at the cell cortex [104]. As a result of their distribution pattern, the PAR proteins determine the segregation of the cytoplasmic RNA-binding proteins MEX-5 and MEX-6 to somatic blastomeres [108]. MEX-5 and MEX-6, in turn, drive the segregation of germline factors to the P blastomeres [109]. Consequently, during the first cell division, the two daughter blastomeres inherit different determinants, which are involved in specifying their distinct fates (anterior/somatic and posterior/germline).

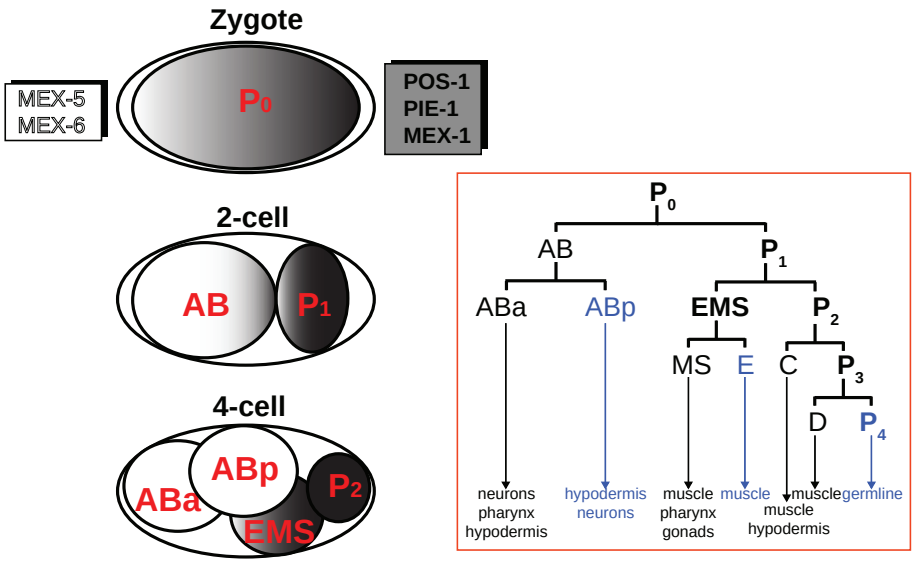


Figure 1.2: Reciprocal gradients of the *C. elegans* CCCH-type TZF proteins. Gradients of the TZF proteins MEX-5, MEX-6, POS-1, PIE-1, and MEX-1 from the 1-4 cell stages of development. MEX-5 and MEX-6 are represented in white, POS-1, PIE-1, and MEX-1 are represented in black. In the red box, the scheme of cell descendants from the P_0 zygote.

1.3.2 The RNA-binding protein MEX-5

MEX-5 and MEX-6 are CCCH-type tandem zinc finger proteins that play a central role in restricting germ plasm factors to the P blastomeres [109]. MEX-5 and MEX-6 share more than 70% of their primary sequence, their contributions to development only partially overlap, however. Disruption of *mex-5* gene, in fact, causes embryonic death with a terminal phenotype that includes proliferation of muscle (Muscle EXcess, MEX). In contrast, deletion of *mex-6* does not affect viability [108, 110].

1.3.3 MEX-5 is required for body axis determination during embryogenesis

Prior to the establishment of PAR polarity, MEX-5 is distributed symmetrically in the zygotic cytoplasm [111]. At this stage, MEX-5 associates with slow-diffusing cytoplasmic complexes that significantly reduce its mobility. Although the nature of these complexes has not been determined yet, it has been reported that they depend on MEX-5 ability to bind RNA [111]. Shortly after fertilization, the posterior kinase PAR-1 phosphorylates MEX-5 on two residues, Serine 404 and Serine 458, resulting in increased MEX-5 diffusion in the posterior cytoplasm. The activity of PAR-1 is opposed by the phosphatase PP2A, symmetrically distributed in the cytoplasm. The combined action of the kinase/phosphatase system generates a gradient in the diffusion rate of MEX-5 along the anterior/posterior axis and leads to the preferential retention of MEX-5 in the anterior [111]. Interestingly, MEX-5 sustains the PAR proteins polarity via a feedback loop with PAR-1, although the mechanism has not been characterized yet [112].

The MEX-5 gradient in the zygote stimulates the redistribution of germ plasm RNA-binding proteins, such as POS-1 and PIE-1, to the posterior cytoplasm, indicating that a primary function of MEX-5 is to restrict germ plasm activities to the P blastomeres [108, 113, 114] (see figure 1.2). In a recent study, Han et al. have provided

useful insights to elucidate the MEX-5-mediated mechanism that lead to POS-1 posterior segregation [115]. They have shown, in fact, that MEX-5 recruits the Polo-like kinase PLK-1 in the anterior cytoplasm. As a consequence, PLK-1 phosphorylates POS-1 and inhibits the formation of stable, slow-diffusing POS-1/RNA complexes in the anterior. The proposed mechanism for POS-1 segregation, therefore, appears similar in many respects to that for MEX-5 segregation.

In support for the a central role of MEX-5 in determining the anterior/posterior pattern of several cell fate determinants, recent studies demonstrated that MEX-5 is also involved in P granules spatial regulation in the embryo. P granules are the *C. elegans* germ granules, ribonucleoprotein organelles specific to the germline. Specifically, MEX-5 has been shown to suppress assembly and stability of P granules in the anterior cytoplasm [116, 117, 118].

1.3.4 The CCCH-type Tandem Zinc Finger domain of MEX-5: comparison with TIS11 proteins

Although belonging to the same CCCH-type TZF family, the TZF domain of MEX-5 diverges from TIS11 proteins in several elements of the primary sequence (see table 1.3). The two zinc fingers of MEX-5 display a different spacing between the three conserved cysteines and the histidine compared to each other and to the TIS11 proteins: $CX_9CX_5CX_3H$ for the N-terminal zinc finger (ZF1), $CX_{10}CX_5CX_3H$ for the C-terminal zinc finger (ZF2) and $CX_8CX_5CX_3H$ for the TIS11 proteins. Moreover, the linker region connecting the two zinc finger is longer in MEX-5 than in TIS11 proteins, 23 residues and 18 residues, respectively.

The (R/K)YKTEL motifs, preceding the first zinc-coordinating cysteine and containing residues that directly contribute to RNA binding [90], is highly conserved in TIS11 proteins across many species. In MEX-5, however, the glutamate residue is

TTP	SRYKTELCRTFSESGR-CRYGAKCQFAHGLGELRQAN---RHP--KYKTELCHKFYLQGR--CPYGSRCHFIHNPSE
MEX-5	PNYKTRLCMMHASGIKPCDMGARCKFAHGLKELRATDAPARYPNNKYKTKLCKNFARGGTGFCPYGLRCEFVHPTDK

Table 1.3: Sequences of the TZF domains of TTP and MEX-5. The two zinc finger motifs (underlined) have the zinc coordinating residues (highlighted in red). Conserved R(K)YKTEL sequences, preceding each finger, are highlighted in blue.

not conserved. The identities of the corresponding amino acids preceding ZF1 and ZF2, are arginine and lysine, respectively.

1.3.5 The RNA-binding activity of MEX-5

Since MEX-5 has been identified as a CCCH-type TZF protein, its RNA-binding activity has been hypothesized on the basis of the homology with the already characterized protein TTP [108]. First studies [110] on the RNA-binding specificity of MEX-5 have confirmed that its TZF domain recognizes, as the human homologs, the ARE of the TNF-α mRNA. Furthermore, it has been shown that MEX-5 binds to RNA molecules containing a single 5′-UUAUUUAUU-3′ sequence, but with an overall affinity that is 5 to 10 folds weaker than TTP.

MEX-5 specificity differs significantly compared to TIS11 proteins. MEX-5 binds to polyuridine sequences with affinity similar to that of TNF-α ARE [110]. Specifically, MEX-5 recognizes sequences containing a tract of 6-8 uridines within an 8-nucleotide window. Sequences with these features are present in the 3′-UTR of 91% of annotated transcripts of *C. elegans* [110]. Therefore, MEX-5 RNA-binding specificity suggests that MEX-5 may act as a broad-spectrum RNA-binding protein to activate maternal mRNA turnover or as an affinity adapter for a more specific proteins, such as PLK-1 [110, 115].

The role of MEX-5 is crucial for the determination of the body axis in the *C. elegans* zygote. Additionally, recent studies unveiled MEX-5 contributions to other

embryogenesis processes, such as P granules formation. However, the lack of knowledge of MEX-5 structure, in particular of its RNA-binding domain, hampers a systematic understanding of the protein functions in the embryo. In this thesis, I present a structural and functional characterization of the TZF domain of MEX-5 that reveals a potential role of MEX-5 during cytokinesis.

1.4 TDP-43 and Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease that causes selective death of motor neurons. The effects of the disease include muscle atrophy, paralysis and eventually death [119]. Although 10% of ALS cases occur on a familial basis, approximately 90% of ALS patients are sporadic [120]. The neuropathology of ALS is characterized by abnormal proteinaceous aggregates in the cytoplasm of degenerating motor neurons [121].

In 2006, TAR DNA-binding protein 43 (TDP43) has been identified as the major disease protein of most forms of ALS [122]. To date, mutations within the gene that encodes TDP43 have been linked with other neurodegenerative diseases, including fronto-temporal lobar degeneration (FTLD) [123], Alzheimer's disease, other tauopathies and Lewy body disorders characterized by α-synuclein inclusions [124].

1.4.1 Protein aggregation and neurodegenerative diseases

A common feature of neurodegenerative diseases, such as Parkinson's disease, Alzheimer's disease, ALS and Huntington's disease, is the presence of insoluble protein aggregates in the brain [125]. These aggregates are classified, based on their cellular localization, as cytoplasmic, nuclear or extracellullar. Although the mechanisms that lead to protein aggregation in diseases are not fully understood, factors that promote aggregation include mutations in disease proteins, genetic alterations that cause increased levels of protein expression or external factors, such as environmental stress or aging [125].

The large intracellular or extracellular accumulations of aggregated proteins, known as inclusion bodies, contain the disease proteins in a fibrillar aggregated form, called amyloid [126]. Proteins in the amyloid state form multimeric assemblies with cross-β structure and are therefore thermodynamically highly stable [127]. Although there is evidence of functional amyloids, for example as natural storage of peptide hormones

[128], in many cases the aggregation of proteins is pathological. Proteinaceous aggregates, in fact, can directly be toxic for the cell, or they can sequester functional proteins within the assemblies [129, 130].

1.4.2 Amyotrophic Lateral Sclerosis and TDP-43 proteinopathies

TDP-43 protein is the major component of the ubiquitinated neuronal cytoplasmic inclusions found in spinal motor neurons in ALS patients and in cortical neurons in FTLD patients [122]. These TDP-43 aggregates are common to 97% of ALS cases, both sporadic or familial [131, 132]. Interestingly, TDP-43 inclusions have not been observed in ALS cases caused by mutations in SOD1 or FUS [131, 133]. Neurodegenerative diseases linked to the deposition of TDP-43 are known as TDP-43 proteinopathies. The hallmark of TDP-43 proteinopathies is the deposition of full-length and fragmented TDP-43 protein as detergent-resistant, ubiquitinated and hyperphosphorylated aggregates in the cytoplasm, with consequent depletion of TDP-43 from the nucleus [134, 122].

In order to establish a direct role for TDP-43 in ALS pathogenesis, several studies have explored the roles of TDP-43 gain or loss of function in the disease. In animal models (mice, *Drosophila* and *C. elegans*), TDP-43 proteinopathy phenotypes are observed upon overexpression of wild-type TDP-43. [135, 136, 137]. Moreover, conditional and partial knockout models of TDP-43 indicate that loss of TDP-43 function induces motor neuron defects, a phenotype reminiscent of the human diseases [138, 139, 140]. Interestingly, selective overexpression or knockdown of TDP-43 in glia and muscle is sufficient to recapitulate ALS-like phenotypes [141, 142]. These results suggest that both gain and loss of TDP-43 function may lead to ALS pathogenesis.

1.4.3 The structure and the function of TDP-43

TDP-43 is a 414-amino acid protein with two RNA recognition motifs (RRM1 and RRM2) and a C-terminal disordered glycine-rich (G-rich) domain. A nuclear localiza-

tion sequence is present at the N-terminus, as well as a nuclear export signal within RRM2. The two RRM domains share significant sequence and structural homology. The structures of the isolated RRM domains (Protein Data Bank entry: 2cqg, 1wf0) show an $\alpha+\beta$ topology for each domain, composed of two repeating $\beta\alpha\beta$ motifs with an extra β strand inserted within the second $\beta\alpha\beta$ motif [143]. Together, these strands form an antiparallel β-sheet across one face of the RRM with the α-helices docked on the opposite face.

The best characterized functions of TDP43 are those concerning its regulation of RNA. The RRM1 domain is critical for TDP-43 RNA-binding activity, specifically to UG repeats [144]. However, association of TDP-43 with sequences not containing UG motifs has been reported [145]. Through its C-terminal G-rich domain, TDP-43 interacts with several heterogenous nuclear ribonucleoproteins (hnRNPs) and consequently TDP43 regulates the splicing of human cystic fibrosis transmembrane conductance regulator [144], survival of motor neuron [146], apolipoprotein A2 (APOA2) [147] and serine/arginine-rich splicing factor 2 (SC35) [148]. TDP-43 also influences mRNA turnover of several genes (cyclin-dependent kinase 6, histone deacetylase 6) [149] and is involved in RNA trafficking [150].

Moreover, TDP43 is a DNA-binding protein, specifically it recognizes single-stranded DNA. TDP43, in fact, has been first identified as a factor recognizing the HIV1 TAR DNA sequence motif [151]. Binding of TDP43 to these DNA sequence results in inhibition of transcription. Additionally, TDP-43 is a potential regulator of acrosomal protein SP10 gene expression during spermatogenesis [152].

1.4.4 Models of disease pathogenesis

TDP-43 proteinopathies are defined by cytoplasmic mislocalization, fragmentation, aggregation, and post-translational modification. Mislocalization of TDP-43 in ALS and FTLD occurs by mean of the deposition of granular and filamentous

macroaggregated protein in the cytoplasm, that results in depletion of TDP-43 from the nucleus [122]. The disease form of TDP-43 corresponds to hyperphosphorylated, ubiquitinated, and proteolytically cleaved C-terminal fragments of $\sim$ 25 kDa [152]. Cleavage of TDP-43 is promoted by C-terminal TDP-43 mutations [153], by cellular stress [154], and proteasomal inhibition [155].

Recent studies have shown that the C-terminal fragments of TDP-43 form insoluble cytoplasmic aggregates [156] and toxic inclusions in cell lines [157, 158]. The C-terminal fragments, in fact, mislocalize to the cytoplasm because of removal of the nuclear localization sequence, and are aggregation-prone due to the prion-like nature of the G-rich domain. Apart from the G-rich region, the truncated RRM2 present in the C-terminal fragments also possess the ability to form fibrils. In particular, it has been shown that β-strands β4 and β5 form two-dimensional sheet-like fibrils, in a way resembling that of the protofilaments of amyloid fibrils [159].

However, several findings highlight that C-terminal fragments may not directly exert toxicity. TDP-43 toxicity, in fact, requires intact RNA binding activity [160]. Moreover, cleavage is not necessary for TDP-43 aggregation, since a cleavage-resistant mutant of TDP-43 is also converted into insoluble species [161]. Several studies in animal models have found that mutant TDP-43 causes toxicity in the absence of visible aggregates [162]. These results suggest that misfolded forms of TDP-43 may play a central role in pathogenesis, directly affecting the protein function and promoting aggregation [134]. Interestingly, Mackness et al. characterized the folding pathways of TDP-43 RRM2 and identified a stable folding intermediate [143]. The study of this intermediate state may be useful to elucidate the mechanisms that link the functional form of TDP43 with its misfolding and aggregation.

Both the RRM2 and the C-terminal region of TDP-43 are required for protein aggregation and toxicity. To date, however, the mechanism of the pathological ag-

gregation of TDP-43 is not completely understood. In this thesis, I characterized the residual structure of a folding intermediate of TDP-43 RRM2, providing insights into the structural changes that might facilitate TDP-43 aggregation.

1.5 Scope of the thesis

This thesis focuses on investigating the relationships between structural disorder and protein activity in RNA-binding proteins. Disorder occurs in many proteins, since lack of structural constraints confers flexibility to the polypeptide chain and facilitate binding to multiple targets. In human, the two paralog CCCH-type tandem zinc finger (TZF) proteins TTP and TIS11d share more than 70% of their sequence. Biophysical characterization of these two proteins shows that both zinc fingers (ZF1 and ZF2) in TIS11d are folded in the absence of RNA. In TTP, however, only ZF1 adopts a stable fold, while ZF2 is unstructured. Additionally, a previous work has demonstrated that the mRNA destabilizing activity of TTP is increased when the partially disordered TZF domain of TTP is replaced with the fully structured domain of TIS11d. In **Chapter 2**, I investigated the interactions that determine the lack of structure in TTP ZF2, and I found that a stacking interaction between the zinc-coordinating histidine and a conserved aromatic side chain is crucial for proper zinc coordination and stability of the finger fold. All together, these results indicate that CCCH-type proteins may have evolved to modulate their activity through their structure.

In *Caenorhabditis elegans* several CCCH-type proteins, with TZF domains homolog of TTP and TIS11d, regulate the first stages of embryogenesis. MEX-5, in particular, determines of the body axis in the zygote and orchestrates the localization of several cell fate determinants. In **Chapter 3**, I characterized the structure of the TZF domain of MEX-5 and I observed that, as in TTP, one of the zinc fingers is unstructured in the absence of RNA. Based on our characterization of the domain, I designed a variant of the protein where both the ZFs are folded in the RNA-free state. I studied the effect of the variant *in vivo*, and we observed that animals homozygous for the mutation are sterile, develop uterine tumors and die by bursting. These results demonstrate that the disorder in the TZF domain of MEX-5 is functional for a still

unknown function of the protein.

Finally, I investigated the role of structural disorder in pathogenic TDP-43 aggregation. TDP-43 is a human RNA-binding protein linked to multiple aspects of RNA biology. Cytoplasmic inclusions, containing full-length or fragments of TDP-43, are the hallmark of many neurodegenerative diseases, particularly ALS and FTLD. Previous studies have shown that the second RNA recognition motif (RRM2) as well as the C-terminal glycine-rich domain are required for aggregation. In **Chapter 4**, I studied the folding intermediate of TDP-43 RRM2 to evaluate its contribution to pathogenic protein aggregation. I found that the ensemble of partially unstructured conformations sampled in this intermediate state exposes to the solvent highly hydrophobic peptide regions. These results provide new insights into the mechanisms underlying TDP-43 proteinopathies.

The last part of this thesis, **Chapter 5**, details how nuclear magnetic resonance spectroscopy and molecular dynamics simulations can be used to characterize the partially unstructured ensemble of conformations typical of disordered proteins. Obtaining structural insights of such diverse conformations, otherwise unaccessible with other biophysical techniques, is crucial to understand the molecular basis of mechanisms underlying pathological mutations or protein functions.

CHAPTER 2

STRUCTURAL BASIS OF THE DISORDER IN THE TANDEM ZINC FINGER DOMAIN OF THE RNA-BINDING PROTEIN TRISTETRAPROLIN

2.1 Abstract

Tristetraprolin (TTP) and TIS11d are two human RNA-binding proteins that belong to the CCCH-type tandem zinc finger family. In the RNA-free state, TIS11d coordinates a zinc ion in each of its two fingers, while TTP coordinates a single zinc ion with the N-terminal zinc finger. We have previously identified three residues, located in the C-terminal half of a short α-helix in the second zinc finger, that control how structured the RNA-binding domain is in these two proteins: Y151, L152, Q153 in TTP and H201, T202, I203 in TIS11d. Here, we have used molecular dynamics, NMR spectroscopy and other biochemical methods to investigate the role of these three residues in the stability of the RNA-binding domain. We found that the intra-helical hydrogen bond formed by the T202 hydroxyl group in the C-terminal zinc finger of TIS11d is necessary to allow for $\pi - \pi$ stacking between the side chains of a conserved phenylalanine and the zinc coordinating histidine. We demonstrated that the lack of this hydrogen bond in TTP is responsible for the reduced zinc affinity of the C-terminal zinc finger.

2.2 Introduction

Tristetraprolin (TTP) is the prototype of the family of CCCH-type zinc finger proteins. In the cell, TTP production is induced by extracellular stimuli such as

insulin, polypeptide growth factor, phorbol esters and mitogens. By controlling the activation of many genes, TTP plays an important role in modulating the inflammatory response [163]. TTP binds to AU-rich elements located in the 3′ untranslated region of many mRNAs, including tumor necrosis factor-α, granulocyte-macrophage colony-stimulating factor, and interleukin-2 [164, 165, 166]. The binding of TTP promotes deadenylation and consequent degradation of these transcripts, thus decreasing the production of these proteins [165, 167, 168]. The RNA-binding domain of TTP is a 70-amino acid fragment that contains the tandem zinc finger domain. The integrity of both zinc fingers is necessary for binding RNA, as any mutation of the CCCH Zn^{2+}-coordinating residues (henceforth defined as Cys^1, Cys^2, Cys^3 and His^4) abolishes binding [169].

There are two other mammalian members in this protein family with TTP-like activity: TIS11b and TIS11d [170]. The RNA-binding domains of these three proteins are highly homologous: the primary sequences of TTP and TIS11d are 71% identical (Fig. 2.10) while those of TIS11d and TIS11b are 91% identical.

The solution structure of the RNA-binding domain of TIS11d bound to the RNA oligonucleotide 5′-UUAUUUAUU-3′ has been solved using NMR spectroscopy [171]. This structure is a novel fold characterized by few secondary structural elements, with the two zinc finger domains, having nearly identical conformations, separated by an extended eighteen-residue linker (Fig. 2.10).

Molecular dynamics (MD) simulations of TIS11d free and bound to RNA have shown that a major structural reorganization occurs in TIS11d when it dissociates from RNA [172, 173]. This reorganization primarily alters the structure of the linker residues, thereby dramatically changing the relative orientations of the two zinc fingers, yet leaving the internal structure of the zinc fingers essentially unperturbed. This structural transition involves the burial of hydrophobic surface area that would otherwise (in the absence of RNA) have an energetically unfavorable exposure to the

solvent. Given the high sequence identity of the RNA binding domains of TTP and TIS11d (Fig. 2.10), these two proteins were expected to have similar structures. Indeed, in the RNA-bound state the secondary structure of TTP, predicted using the backbone and C_β chemical shifts using the δ2D [174] and SPARTA+ [175] programs, is the same as that of TIS11d [176]. In the absence of RNA, however, only the N-terminal zinc finger (ZF1) of TTP adopts a stable fold while the C-terminal zinc finger (ZF2) does not stably bind Zn^{2+} [176, 177, 178]. The effect of a fully folded RNA-binding domain on the cellular activity of TTP has been determined using a luciferase reporter assay, where luciferase was placed under the control of the TNF-α 3′ UTR [176]. Lower reporter activity was observed when the partially disordered RNA-binding domain of TTP was replaced with the fully structured domain of TIS11d, indicating that increased structure is associated with higher RNA-degradation activity [176]. This result showed that folding of the RNA-binding domain is tightly coupled with the activity of TTP and TIS11d in the cell.

We have previously shown [176] that the stability of ZF2 is determined by the identity of three residues, located in the C-terminal half of the α-helix of ZF2: Y151, L152, Q153 in TTP and H201, T202, I203 in TIS11d. Here, we investigate the following question: how do the residues located at the C-terminal half of the α-helix determine the affinity of ZF2 for Zn^{2+} and consequently the folding and stability of ZF2? To address this question we simulated TIS11d and a homology model of TTP in solution using molecular dynamics. Analysis of the resulting trajectories points to specific key interactions that stabilize the structure of ZF2 in TIS11d but that are absent in TTP. To validate the findings from simulation, we tested the role of these specific interactions using mutagenesis, NMR and CD spectroscopy experiments.

2.3 Results and Discussion

2.3.1 ZF2 in TTP does not maintain Zn^{2+}-coordination in the MD simulations

MD simulations of TIS11d in the free state show that the linker region of TIS11d (Fig. 2.10) is flexible in solution, while the structures of both ZF1 and ZF2 are maintained throughout all trajectories, in agreement with the experimental spectroscopic data [171, 176, 172, 173]. The homology model of the RNA-binding domain of TTP used in this study was based upon the solution structure of TIS11d. For this reason, the initial structure of TTP used in the MD simulations has both zinc fingers folded and coordinating Zn^{2+} ions. Experimental evidence indicates, however, that only ZF1 of TTP can stably bind Zn^{2+} in the RNA-free state [176, 177, 178]. Consistent with this evidence, one of the three MD trajectories of TTP shows the loss of Zn^{2+}-coordination at ZF2, Fig. 2.1, through a series of events that are described in detail below. The remaining two trajectories exhibit the earliest events that promote this loss (Figs. 2.14 and 2.15). To characterize the Zn^{2+}-coordination of each finger, the six angles and the four distances between the Zn^{2+} cation and the Zn^{2+}-coordinating atoms (S from the cysteine residues, and N_ε from the histidine) were monitored as a function of time, Fig. 2.1.

To determine the order of events that lead to the loss of Zn^{2+}-coordination at ZF2, the structure, intramolecular fluctuations and overall dynamics were analyzed from the trajectory where this loss was exhibited, as described in the following sections.

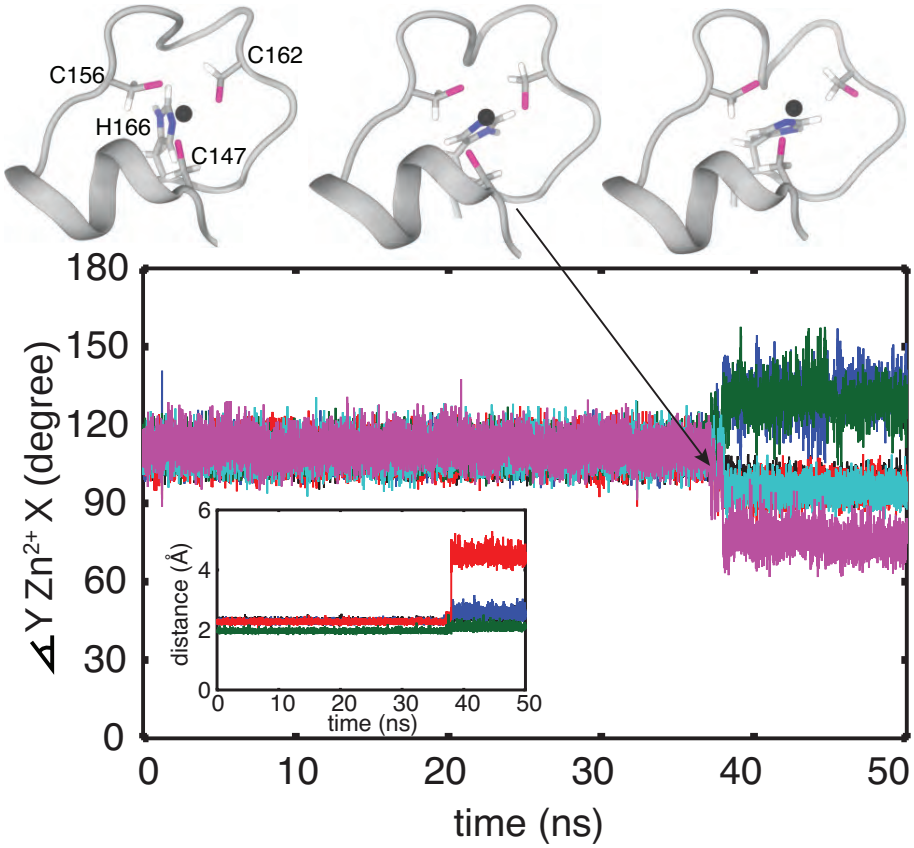


Figure 2.1: Unfolding of ZF2 during MD simulation of TTP. Conformations of the unfolding C-terminal zinc finger of TTP, sampled from an MD trajectory, are shown at top and correspond to t =37.130 ns (top, left), 37.145 ns (top, center) and 38.420 ns (top, right). Below, the geometry of the zinc coordination in the C-terminal zinc finger of TTP is monitored. In the main figure, the angles between the zinc ion and the zinc coordinating atoms are shown for the first 50 ns: $\measuredangle S^{C147}$-Zn^{2+}-S^{C156} in black, $\measuredangle S^{C147}$-Zn^{2+}-S^{C162} in blue, $\measuredangle S^{C147}$-Zn^{2+}-N_{ε}^{H166} in red, $\measuredangle S^{C156}$-Zn^{2+}-S^{C162} in green, $\measuredangle S^{C156}$-Zn^{2+}-N_{ε}^{H166} in cyan and $\measuredangle S^{C162}$-Zn^{2+}-N_{ε}^{H166} in magenta. In the inset, the distances between the zinc ion and the zinc coordinating atoms are shown: S^{C147}-Zn^{2+} in black, S^{C156}-Zn^{2+} in blue, S^{C162}-Zn^{2+} in red and N_{ε}^{H166}-Zn^{2+} in green.

2.3.2 The Zn^{2+}-coordinating histidine in ZF2 of TTP samples two conformations

Comparison of the trajectories collected for TTP and TIS11d reveals important differences in the structures of ZF2 in the two proteins. In particular, analysis of the dihedral angles χ_1 (defined by the following atoms: N, C_α, C_β, C_γ) and χ_2 (defined by the following atoms: C_α, C_β, C_γ and N_{δ_1}) for the Zn^{2+}-coordinating residues shows that in TTP the Zn^{2+}-coordinating histidine (i.e. His^4) in ZF2 (H166 in TTP) samples two different states characterized by different orientations of the aromatic side chain: $\chi_2 = 180°$ and $\chi_2 = 90°$ (see Fig. 2.2B). The His^4 residues in ZF1 of TTP and in both ZFs of TIS11d mostly sample the conformation characterized by $\chi_2 = 180°$, and visit the second conformation, characterized by $\chi_2 = 90°$, very infrequently, with $\approx 1\%$ probability (see Fig. 2.2A,C,D).

Rotation of the imidazole ring of H166, Fig. 2.3A, leads to steric hindrance with the side chain of C162, the third Zn^{2+}-coordinating residue in ZF2, Fig. 2.3B. As a consequence, C162 separates from both H166 and the Zn^{2+} cation, thereby disrupting proper Zn^{2+}-coordination, Fig. 2.3C. Thus, one of the two orientations of the H166 side chain, $\chi_2 = 90°$, is incompatible with Zn^{2+}-binding.

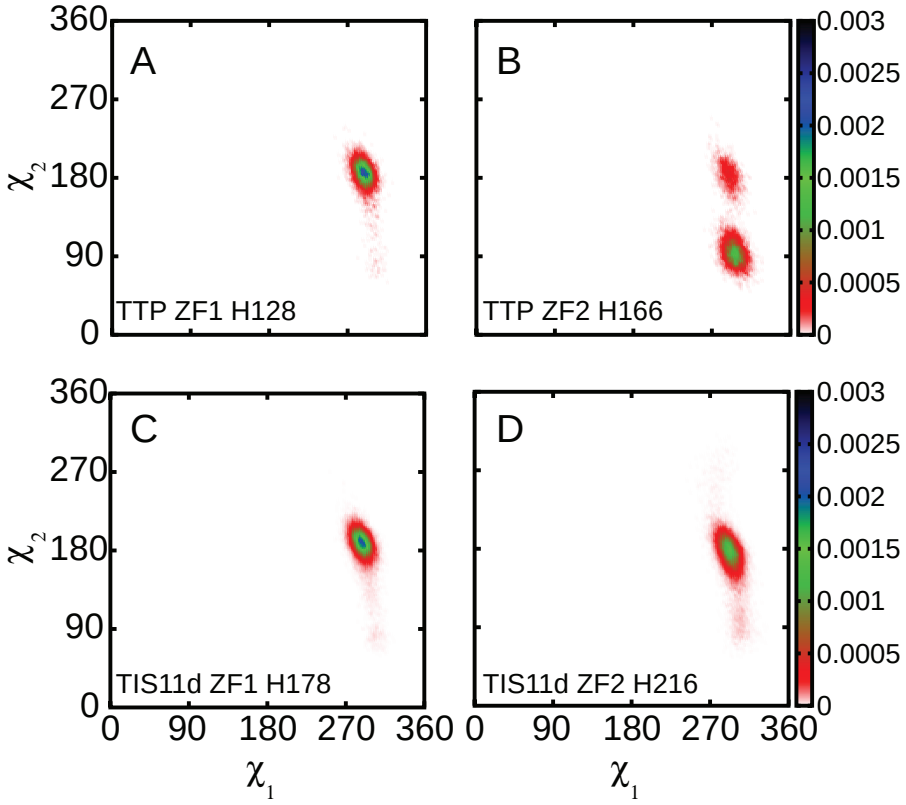


Figure 2.2: His[4] of ZF2 populates two rotameric conformations in TTP but not in TIS11d. Probability density distribution of the dihedral angles χ_1 and χ_2 for the side chains of H128 (A) and H166 (B) of TTP and H178 (C) and H216 (D) of TIS11d. Data are taken from the 100 ns MD trajectory of TTP where loss of Zn^{2+}-coordination was observed and from the six 100 ns MD trajectories of TIS11d. The color bars show the values of the probability density calculated for χ_1 and χ_2 as the number of counts normalized by the total number of observations and by the area of each bin.

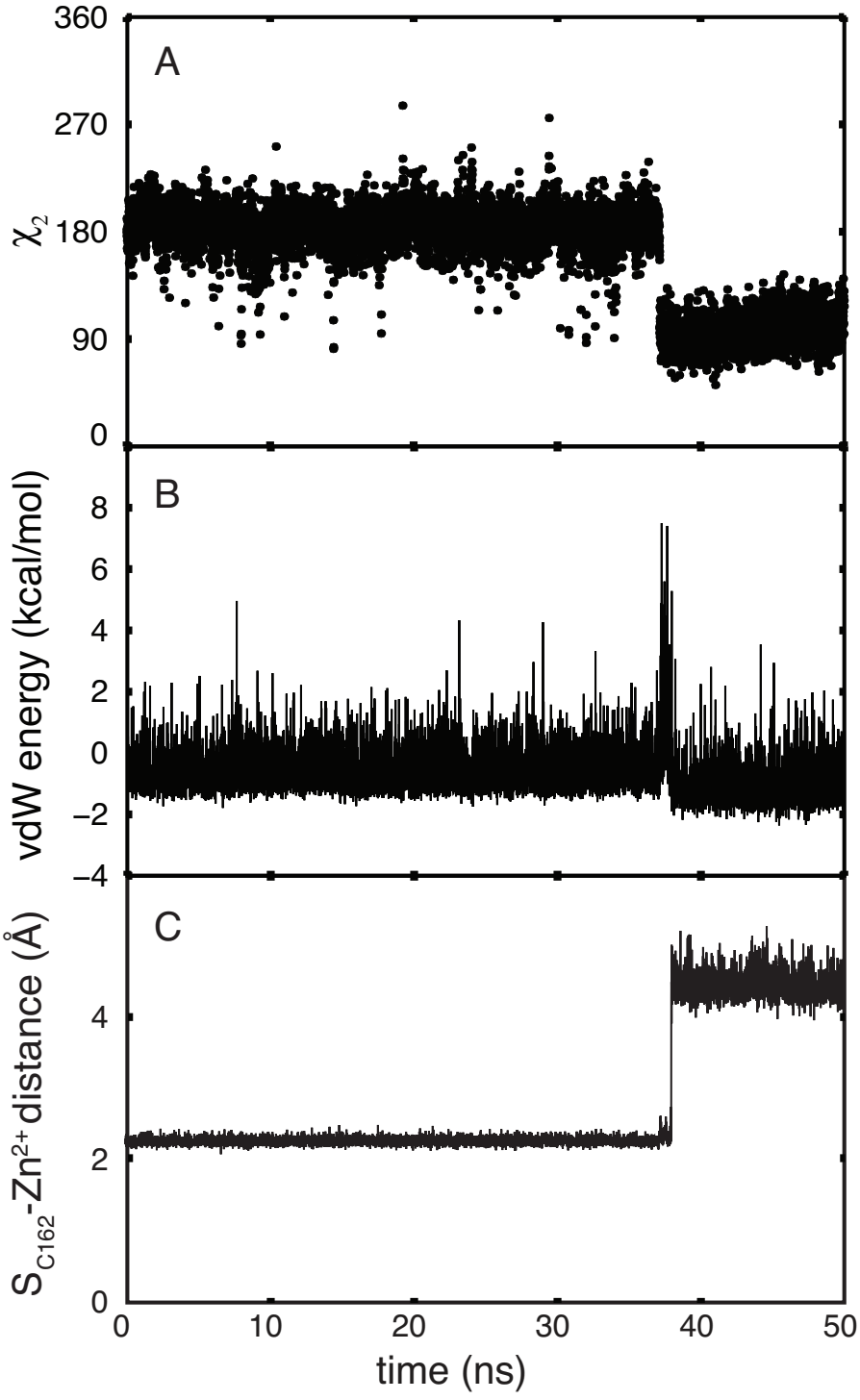


Figure 2.3: Side chain rotation of H166 causes a steric clash with C162. The χ_2 dihedral angle of H166 (A), van der Waals interaction energy between residues C162 and H166 (B) and S^{C162}-Zn^{2+} distance (C) are shown as functions of time. Data are shown for the first half of the unfolding MD trajectory of TTP (100 ns in total).

2.3.3 The stacking of histidine and phenylalanine aromatic rings stabilizes Zn^{2+}-binding in ZF2

The interaction of H166 with F150, in TTP, and of H216 with F200, in TIS11d, stabilizes the His^4 side chain in the Zn^{2+}-binding conformation, characterized by $\chi_2 = 180°$. For TIS11d, both ZF1 and ZF2 exhibit stacking of aromatic residues throughout our simulations. The aromatic side chain of His^4 stacks with that of a conserved phenylalanine located in the α-helix, three residues after the first Zn^{2+}-coordinating cystein, Cys^1+3 (see Fig. 2.4 and Fig. 2.16). For TTP however, while this aromatic stacking is consistently observed in ZF1 (see Fig. 2.16), it is observed in ZF2 only in the first part of the trajectory that preceeds to the loss of Zn^{2+}-coordination (see Fig. 2.4). When this stacking interaction between H166 and F150 is lost, H166 preferentially samples the rotameric configuration characterized by $\chi_2 = 90°$ (Fig. 2.5), resulting in the loss of Zn^{2+}-coordination from ZF2.

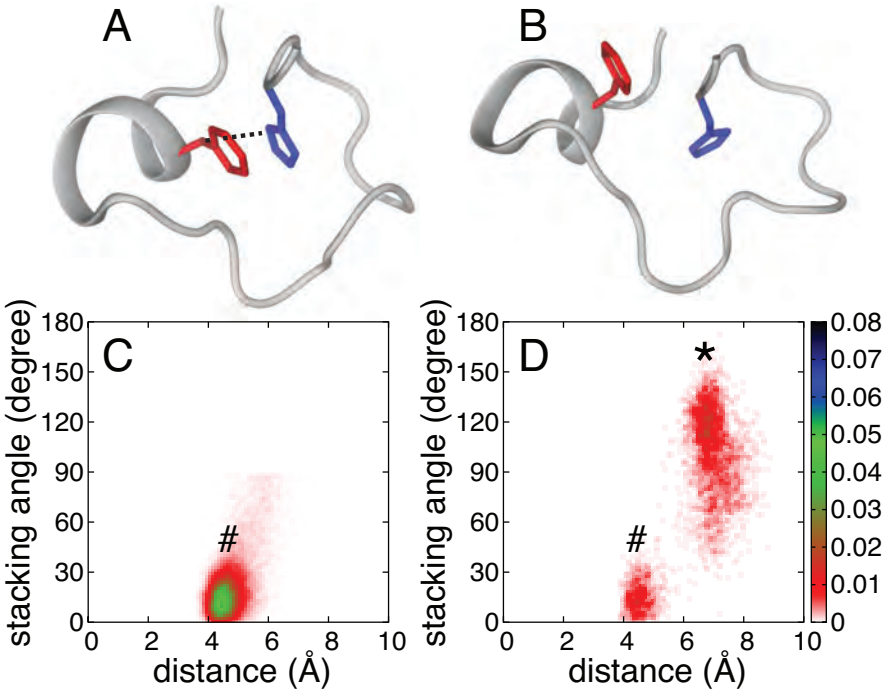


Figure 2.4: Aromatic side chain stacking of phenylalanine and histidine in ZF2. Representative ZF2 structures for TTP showing F150 (red) and H166 (blue) side chains stacked (A) and not stacked (B). Probability density distributions of the stacking angle and distance between the aromatic rings are shown for F200 and H216 in TIS11d (C) and for F150 and H166 in TTP (D). In the figure, conformations where the phenylalanine and histidine side chains are stacked (as shown in A) and not stacked (as shown in B) are represented by a # or a $*$, respectively. The stacking angle was calculated as the angle between the normals of the two aromatic rings (the planes for the side chains are defined by atoms C_{δ_2}, C_{ε_1} and N_{ε_2} for histidine and C_{ζ}, C_{ε_2}, C_{δ_2} for phenylalanine). The distance between the aromatic rings was calculated as the distance between the centers of mass for the heavy atoms of the two side chains. The color bars show the values of the probability density calculated for the stacking angle and distance as the number of counts normalized by the total number of observations and by the area of each bin. Configurations and distributions were extracted from the unfolding MD trajectory for TTP (100 ns) and from six 100 ns MD trajectories for TIS11d.

2.3.4 The stacking interaction between histidine and phenylalanine occurs if the α-helix axis is bent

As described above, the phenylalanine residue that interacts with His^4 is located in the α-helix that separates Cys^1 and Cys^2 in each zinc finger. Structural alignment of the α-helices of ZF2 of TIS11d and TTP reveals a difference in their conformations (see Fig. 2.6) that is also evident from the difference in the backbone dihedral angles of the three residues located in the C-terminal half of the α-helix (see Fig. 2.17). In TIS11d, the α-helix axis is bent due to the presence of a hydrogen bond between the side chain hydroxyl of T202 (the fifth residue in the α-helix, Cys^1+5 position) and the backbone acyl group of R198 (Cys^1+1 position). This hydrogen bond is present in all trajectories collected for TIS11d. The equivalent hydrogen bond in ZF1, between the O_γ of S165, Cys^1+6, and the acyl oxygen of P161 in TIS11d and between the corresponding residues S115 and T111 in TTP, is always observed in all the trajectory collected for TIS11d and TTP. A hydrogen bond between the O_γ atom of a serine or threonine residue with the acyl oxygen of the $i-3$ or $i-4$ residue has been previously observed to induce a bend in the α-helix axis of $\approx 3-4^\circ$, particularly in transmembrane α-helices [179].

Because of the lack of the corresponding hydrogen bond in TTP (between the side chain of Cys^1+5 and the backbone of Cys^1+1), the bend in the α-helix is not stabilized as it is in TIS11d. Fig. 2.6 shows the straightening of the α-helix axis occurring in TTP at $\approx$ 37 ns. This rearrangement of the α-helix results in a displacement of the phenylalanine side chain located at the middle of the helix, Cys^1+3. In this new position, the phenylalanine side chain is not positioned to stack against the imidazole ring of H166 as well as the corresponding stacking interaction in ZF2 of TIS11d (see Figs. 2.4 and 2.6). This helix reorganization is the first step in the zinc finger destabilization that eventually leads to the loss of Zn^{2+}-coordination in TTP.

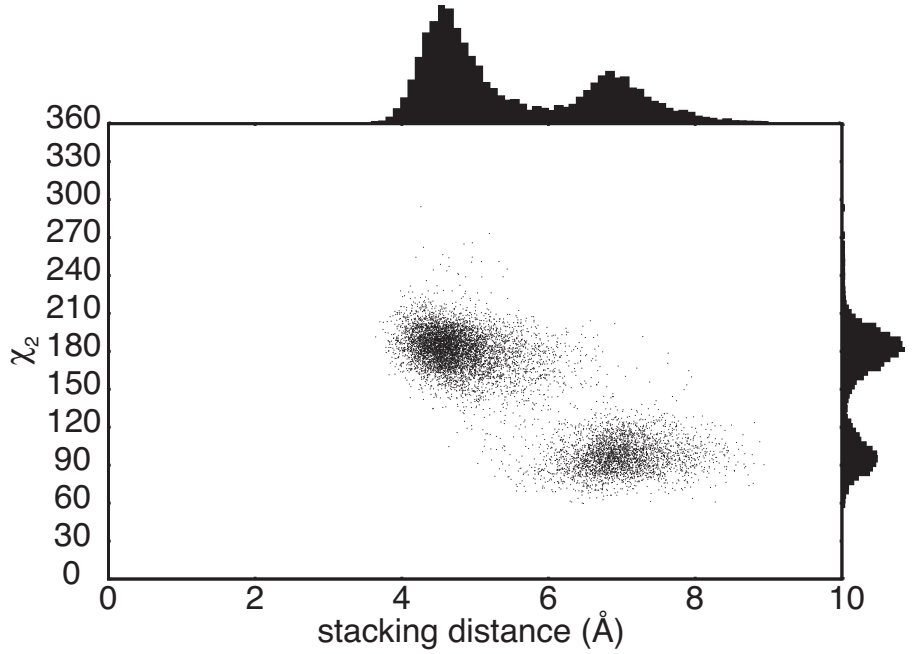


Figure 2.5: H166 in TTP populates the rotameric conformation with $\chi_2 = 90°$ upon loss of stacking with F150. Scatter plot of the stacking distance between the aromatic ring of F150 and H166 of TTP and the χ_2 dihedral angle of H166. Probability distributions of the stacking distance and χ_2 dihedral are shown on the axes. The distance between the aromatic rings was calculated as the distance between the centers of mass for the heavy atoms of the two side chains. Data were extracted from the three MD trajectory for TTP.

2.3.5 MD simulations of the TIS11d T202L mutant show decreased stacking between the side chains of H216 and F200

Comparison of the MD simulations of TTP and TIS11d highlighted the role of the hydrogen bond between the backbone acyl oxygen of R198 and the side chain of T202 of TIS11d in stabilizing the bend in the α-helix. This bent conformation of the α-helix places the aromatic ring of F200 in a position suitable for stacking against H216, thus mantaining H216 in the rotameric state that allows Zn^{2+}-binding.

To further assess this mechanism of stabilization of ZF2, we collected six 100 ns MD trajectories of a mutant of TIS11d where T202 was mutated into leucine, the equivalent residue found in TTP. Due to the absence of an hydroxyl group on the side chain of the leucine, this mutant cannot form the hydrogen bond. For this reason the α-helix located in ZF2 is not stabilized in a bent conformation, and the stacking interaction between F200 and H216 is not as stable as that in the wild type protein. This relative instability results in H216 sampling both rotameric side chain conformations, one that allows ZF2 to stably coordinate Zn^{2+} ($\chi_2 = 180°$) and one that does not ($\chi_2 = 90°$), as shown in Figs. 2.18 and 2.19. The conformation that is incompatible with Zn^{2+}-binding is sampled by H216 with higher probability (9.6%) in the mutant than in the wild type protein (4.7%), due to the lack of a stacking interaction with F200 (Figs. 2.18 and 2.19).

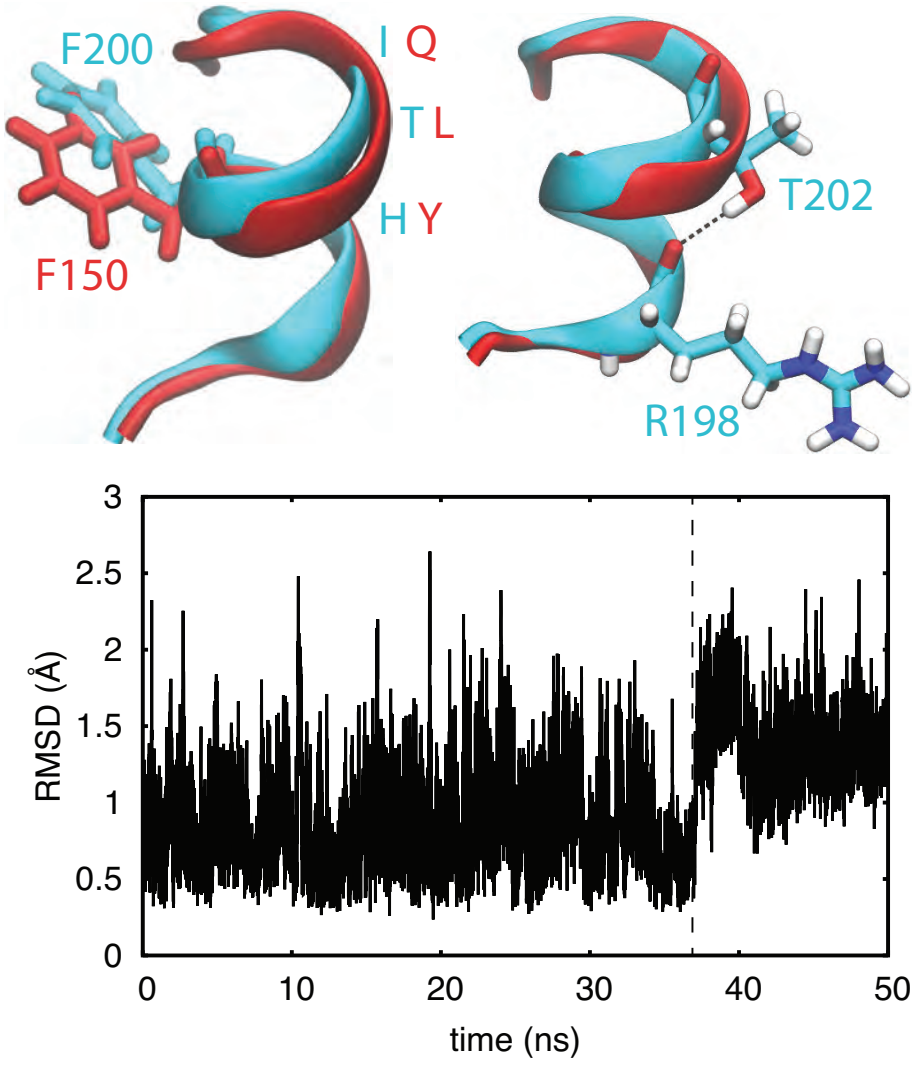


Figure 2.6: The hydrogen bond between T202 and R198 stabilizes a bend in the axis of the α-helix. Top: Structures of the C-terminal zinc finger of TTP (red) and TIS11d (cyan). The orientations of the side chain of F150 (TTP, red) and F200 (TIS11d, cyan) are shown on the left. The hydrogen bond between O_{γ}^{T202} and O^{R198}, depicted as black dashed line, is shown on the right. Oxygen atoms are depicted in red, nitrogen in blue and hydrogen in white. Bottom: The root-mean-square deviation of the backbone of the α-helix in TTP (residues 147 to 153) is shown as a function of time. The dashed line indicates the change in the α-helix conformation that causes a displacement of the F150 side chain to a position where it does not stack against H166. Data are shown for the first half of the unfolding MD trajectory of TTP (100 ns in total)

2.3.6 Mutation of threonine 202 to leucine is sufficient to destabilize the structure of ZF2 in TIS11d

Our simulations of TTP and TIS11d indicate that the structural difference observed between ZF2 of TTP and that of TIS11d arises from the different amino acid composition of the α-helix, in agreement with reported experimental observations [176]. In particular, our simulations reveal that the conformation of the α-helix is different in the two proteins. In TIS11d, the hydrogen bond between the side chain of T202 and the backbone of R198 causes a bend in the α-helix axis, thereby positioning the aromatic ring of F200 to stack against the imidazole ring of H216. Stacking of the aromatic side chains of F200 and H216 stabilizes H216 in the rotameric conformation ($\chi_2 = 180°$) that stabilizes Zn^{2+}-binding in ZF2 (see Fig. 2.3). In TTP, the homologous residue to T202 in TIS11d is L152, whose side chain cannot form a hydrogen bond with the backbone of residue 148. For this reason, the α-helix axis is not bent in TTP as it is in TIS11d, resulting in a lower probability for the F150 side chain of stacking against that of H166 (compare Figs. 2.4 A and B). Thus, F150 does not stabilize the H166 side chain in the conformation with $\chi_2 = 180°$ (Fig. 2.5), that maintains Zn^{2+}-coordination in ZF2 (Fig. 2.3).

To validate the role of T202 in stabilizing the structure of ZF2 based on our MD results, we mutated threonine 202 to leucine in TIS11d and used NMR and CD spectroscopy to characterize the structure of the mutant protein. We found that in T202L mutant of TIS11d, ZF2 is less structured than in the wild type protein, Figs. 2.7 A and B. Cross-peaks from ZF1 and ZF2 are present in the ^{15}N-^{1}H HSQC spectrum of the T202L mutant TIS11d, however cross-peaks corresponding to the residues in the linker and in the α-helix of ZF2 are broadened beyond detection, Fig. 2.7A. Cross-peaks from ZF2 show the largest chemical shift differences from the wild type and have lower intensities than cross-peaks from ZF1, Fig. 2.20. These results suggest that ZF2 is more flexible and that the structure of ZF2 is affected by the single point

mutation. The CD spectrum of the RNA-binding domain of TIS11d T202L indicates that this protein is less structured than the wild type TIS11d, and more similar to TTP, Fig. 2.7B.

Zn^{2+} titrations of TIS11d T202L monitored by NMR spectroscopy show that only ZF1 stably binds the metal cation. The titration endpoint occurred at a [Zn^{2+}] / [TIS11d T202L] ratio of 1:1, and further addition of Zn^{2+} resulted in no changes in the cross-peaks intensity or in their position, Figs. 2.21 and 2.22. Cross peaks from ZF2 have low signal-to-noise ratios, indicating that a small fraction of ZF2 is folded, and that ZF2 binds Zn^{2+} with lower affinity than ZF1.

In agreement with what is observed for TTP, addition of RNA stabilizes the structure of ZF2 in TIS11d T202L, as indicated by the presence of cross-peaks from ZF2 in the ^{15}N-^{1}H HSQC spectrum of the T202L mutant TIS11d that were missing in the free state, Fig. 2.23. All together these results indicate that mutation of threonine 202 to leucine decreases the affinity of ZF2 for Zn^{2+} and destabilizes its structure.

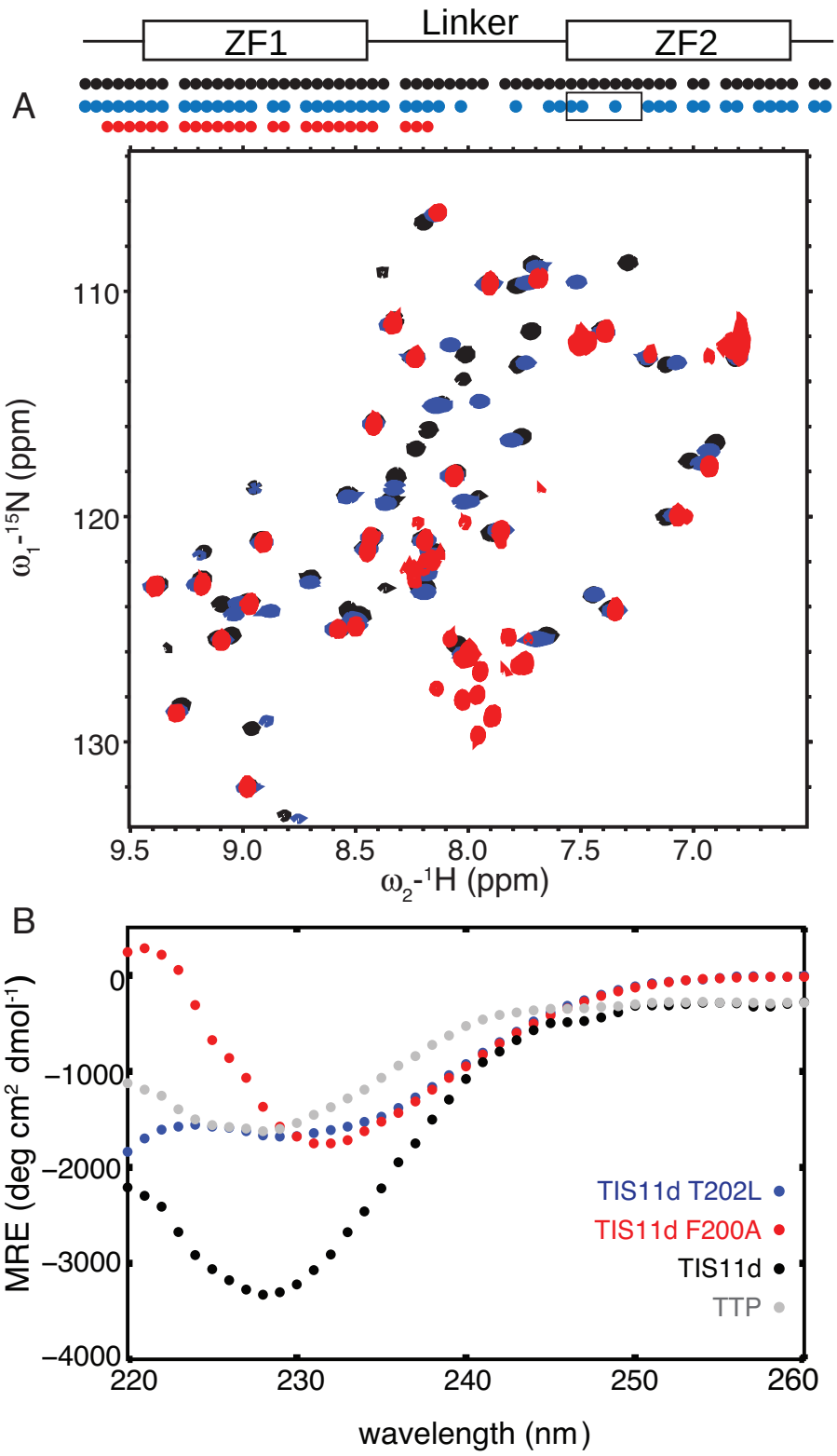

Linker
ZF1
ZF2
A
110
120
130
ω1-15N (ppm)
9.5
9.0
8.5
8.0
7.5
7.0
ω2-1H (ppm)
B
0
−1000
−2000
−3000
−4000
MRE (deg cm2 dmol-1)
220
230
240
250
260
wavelength (nm)
TIS11d T202L
TIS11d F200A
TIS11d
TTP

Figure 2.7: Structural studies of T202L and F200A mutant of TIS11d. ^{15}N-^{1}H HSQC spectra (A) of TIS11d wild type (black), T202L (blue) and F200A (red) mutants of TIS11d. On top, a schematic representation of the RNA-binding domain depicts the ZFs as rectangles and the linker region as a line. The circles indicate residues along the primary sequence with a cross-peak in the ^{15}N-^{1}H HSQC spectrum. Black box indicates the α-helix in ZF2. The F200A mutant protein is less stable than the wild type protein as indicated by the presence of degradation peaks in the region between 7.8 and 8.2 ppm in the ^{1}H dimension and 125 and 130 ppm in the ^{15}N dimension. Circular dichroism spectra (B) of TIS11d wild type (black), T202L (blue) and F200A (red) mutants of TIS11d and TTP (gray).

2.3.7 Mutation of leucine 152 to threonine is sufficient to stabilize the structure of ZF2 in TTP

Following the same rationale discussed in the previous section that lead to the mutation of threonine 202 to leucine in TIS11d, we made the equivalent mutation to TTP, L152T. The introduction of a threonine in the middle of the α-helix allows for the formation of a hydrogen bond with the acyl oxygen of the forth preceeding residue, as observed in TIS11d. Such a hydrogen bond introduces a bend in the α-helix that supports the stacking of phenylalanine located on the opposite side of the helix with the Zn^{2+}-coordinating histidine. For this reason, we expected ZF2 of this mutant to have higher affinity for Zn^{2+} than the wild type protein. As shown by NMR and CD spectroscopy, this mutant stably binds Zn^{2+} with both zinc fingers (Fig. 2.8). The ^{15}N-^{1}H HSQC spectrum of TTP L152T contains cross-peaks from both ZF1 and ZF2 (Fig. 2.8A) and the CD spectrum closely resembles that of TIS11d (Fig. 2.8B). Combined these results indicate that both zinc fingers are properly folded in this mutant protein.

Taken together TIS11d T202L mutant and the corresponding TTP L152T mutant confirm the role of the intra-helical hydrogen bond formed by the side chain of the threonine in stabilizing the $\pi - \pi$ interaction, essential for the stability of ZF2.

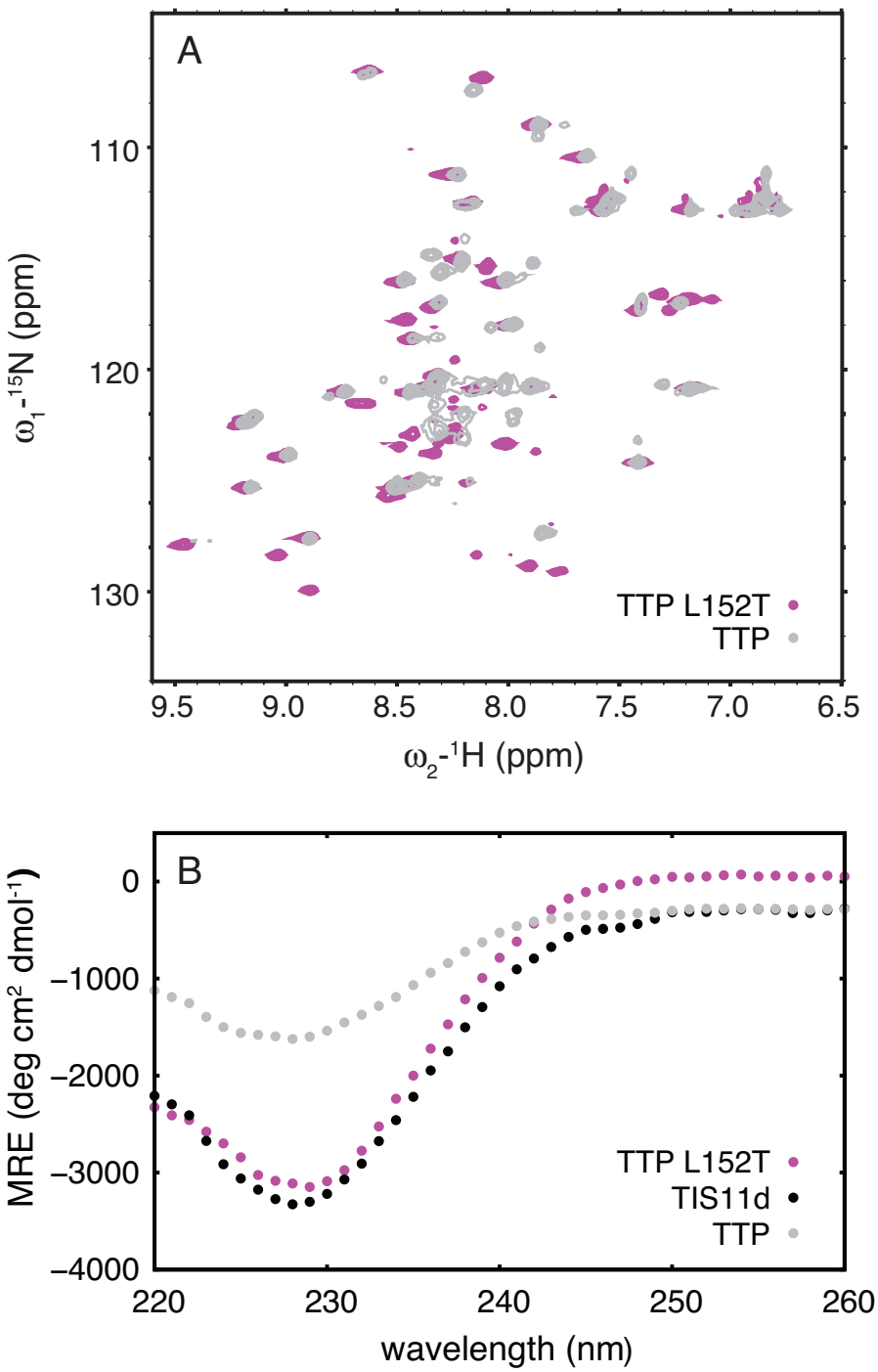


Figure 2.8: Structural studies of L152T mutant of TTP. ^{15}N-^{1}H HSQC spectra (A) of TTP (gray) and L152T mutant of TTP (magenta). Circular dichroism spectra (B) of TIS11d wild type (black) and TTP wild type (gray) and L152T mutant (magenta).

2.3.8 ZF2 is unfolded in the F200A mutant of TIS11d

To validate the importance in TIS11d of the aromatic stacking interaction between F200, located in the middle of the α-helix, and H216 in stabilizing the structure of the ZF2, we mutated F200 to alanine. The ^{15}N-^{1}H HSQC spectrum of the F200A mutant TIS11d is missing all cross-peaks from ZF2, indicating that ZF2 samples multiple states and, as a consequence, cross-peaks are broadened beyond detection (see Fig. 2.7A). Cross-peaks from ZF2 are observed in the ^{15}N-^{1}H HSQC spectrum of wild type TIS11d, where both ZFs are folded, but are missing in the ^{15}N-^{1}H HSQC spectrum of TTP, where only ZF1 is folded. Missing cross-peaks from ZF2 are an indication of lack of ZF2 structure. To prove it, we compared the spectrum of F200A with that of a mutant protein of TIS11d where ZF2 is known to be unfolded, C212S [176] (Fig. 2.24). In TIS11d C212S, Cys^3 in ZF2, is mutated to serine thus, by removal of one of the essential Zn^{2+}-coordinating residues, zinc binding is abrogated and the structure of ZF2 destabilized [176].

We used CD spectroscopy to characterize the structure of the mutant protein. Fig. 2.7B shows that the RNA-binding domain of F200A is less structured than that of wild type TIS11d, and more similar to that of TTP. The shift in minimum of the CD spectrum is due to the loss of the aromatic interactions in the F200A mutant. These experimental data demonstrate that F200 is essential in stabilizing the structure of ZF2. It is important to note that F200 is located on the protein surface, it is partially solvent exposed and there are no hydrophobic side chains within a 5 Å radius (see Fig. 2.25). The destabilization of ZF2 observed in the F200A mutant, therefore, is not due to the disruption of a hydrophobic core within the protein but only to the disruption of the aromatic stacking with H216.

While RNA-binding by TTP is sufficient to stabilize the structure of ZF2 and bind a Zn^{2+} cation [176], RNA-binding is not sufficient to stabilize the structure of ZF2 in the F200A mutant of TIS11d. The ^{15}N-^{1}H HSQC spectrum of F200A shows

minor changes upon addition of RNA: small shifts are present in the position of the peaks from ZF1 and no additional peaks appear from ZF2 residues (see Fig. 2.26). The addition of RNA, therefore, does not stabilize the structure of ZF2 and this mutant protein does not bind RNA with high affinity, K_d ¡ 1 μM, as confirmed by a fluorescence polarization binding assay. These data show that a phenylalanine located three residues after the first Zn^{2+}-coordinating cysteine (i.e. Cys^1), in the middle of the α-helix in TIS11d, is essential for stabilizing the structure of the zinc finger and for increasing the binding affinity of the zinc finger for Zn^{2+}.

2.3.9 Sequence alignment

Analysis of the primary sequences of 14,851 zinc finger domains from the tz-CCCH family obtained from PFAM (PF00642) [180] shows that the aromatic character of third residue after the first Zn^{2+}-coordinating cysteine (Cys^1+3) is highly conserved (see Fig. 2.9A). In addition, we have shown that a bend in the α-helix axis facilitates the stacking of the phenylalanine at the Cys^1+3 position with His^4. This bend is stabilized by the formation of a hydrogen bond between the O_γ of either a threonine or a serine residue, located either at Cys^1+5 or Cys^1+6, and the acyl oxygen of the preceding fourth residue. Primary sequence analysis of the tz-CCCH family [180] shows that a residue containing an O_γ, although not fully conserved, is present with high frequency at either Cys^1+5 or Cys^1+6 (see Fig. 2.9B), with threonine being the most abundant amino acid (46%). A charged or polar residue is often observed at either of these positions with Arg, Lys and Glu having the highest probabilities after threonine. The probability of finding an Asp is much lower than that of Glu, suggesting that the side chain length of the charged amino acid is important.

Primary sequence analysis indicates that the stacking of an aromatic side chain with the Zn^{2+}-coordinating histidine is the strategy adopted by a large majority (94.2%) of CCCH-type zinc finger domains to coordinate Zn^{2+} with high affinity

(see Fig. 2.9A). Stabilization of the stacking interactions of the aromatic side chains, however, is likely achieved in different ways by different zinc fingers in this family. We have shown that in TIS11d, both ZF1 and ZF2 use an intra-helical hydrogen bond to stabilize the stacking interactions between the phenylalanine in the middle of the α-helix and His^4. In TTP, however, only ZF1 adopts this strategy to stabilize its structure, while ZF2 binds Zn^{2+} with low affinity in the absence of RNA and is partially unstructured in solution. Our previous studies indicate that TTP has recently evolved to modulate its activity through its folded stability [176]. The CCCH-type tandem zinc finger domain contains few secondary structural elements and thus has a relatively low thermodynamic stability. A single point mutation, therefore, can easily destabilize the fold and shift the equilibrium towards a disordered zinc finger state. For this reason, we expect proteins from this family to have evolved to use an unfolded-to-folded transition to regulate their activity in the cell.

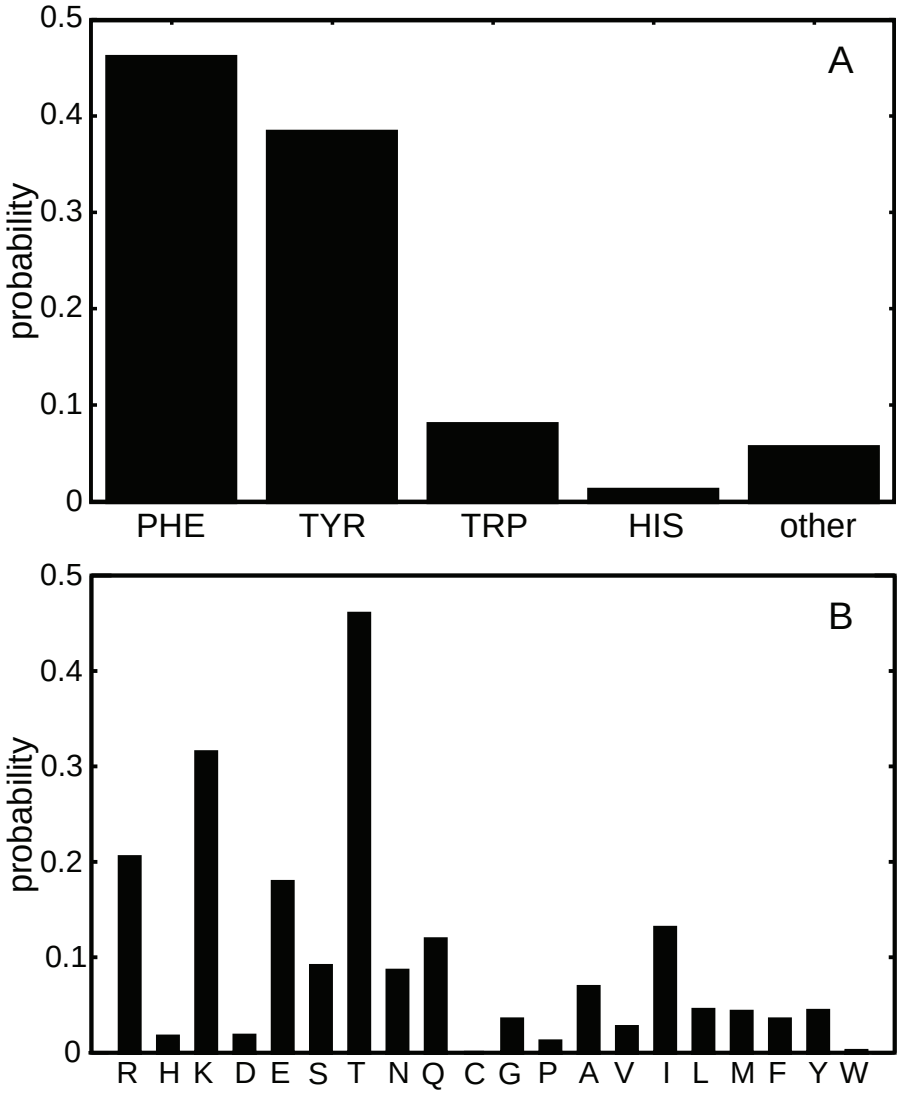


Figure 2.9: Sequence analysis of the tz-CCCH family (PFAM ID: PF00642). A: Probability of finding an aromatic or not aromatic residue at position Cys^1+3. B: Probability of finding a given amino acid at position Cys^1+5 or Cys^1+6.

2.4 Conclusions

In this work, we investigated the interactions occurring in the RNA-binding domains of TTP and TIS11d that stabilize their coordination of zinc ions. Using molecular dynamics we were able to observe the loss of structure of the C-terminal zinc finger of TTP and characterize the events that underly this loss. We found that, in the C-terminal zinc finger of TTP, the zinc coordinating histidine, H166, populates two rotameric states. The rotamers correspond to the χ_2 angle of the side chain centered at 180° or at 90°, respectively: the first conformation allows the correct tetrahedral geometry between the three cysteines, the histidine and the zinc ion; the latter causes a steric clash between the side chain of H166 and C162 that results in the disruption of the zinc binding site. We found that when the rotameric state of H166 has $\chi_2 = 180°$, a $\pi - \pi$ interaction is present between the side chains of H166 and F150; when $\chi_2 = 90°$, no such interaction is present. When the stacking between the two aromatic moieties is formed, the side chain of histidine residue is kept in the $\chi_2 = 180°$ conformation. In TTP, this stacking interaction is only marginally stable, however.

In TIS11d, the rotamer of H216 with $\chi_2 = 180°$ is stabilized by the $\pi - \pi$ interaction with the side chain of F200. As in TTP, the phenylalanine is in the center of the short α-helix spanning six residues and starting at the first cysteine. An intra-helix hydrogen bond between the hydroxyl group in the side chain of T202 and the acyl oxygen of R198 restrains the α-helix in a conformation that allows the side chain of F200 to stack against the imidazole ring of H216. Although TTP and TIS11d share the majority of their primary sequence in the RNA-binding domain, the residues that form the α-helix are not conserved. The residue corresponding to T202 in TIS11d is L152 in TTP. Leucine side chains are unable to form hydrogen bonds and hence in TTP the phenylalanine is not kept in proximity of the imidazole ring of H156 in a conformation that can stack against the side chain of H156. MD simulations of

TIS11d T202L mutant support the importance of this hydrogen bond in stabilizing the Zn^{2+}-coordination: in the mutant protein, which cannot form the hydrogen bond, the H216 rotamer with $\chi_2 = 90^\circ$ is more populated than in the wild type TIS11d.

The mechanism of stabilization of ZF2 that we proposed based on the MD simulations was experimentally validated using mutagenesis. First, we constructed two mutant proteins, TIS11d T202L and TTP L152T, to probe the role of the hydrogen bond in stabilizing the $\pi - \pi$ interaction observed between the side chains of the Zn^{2+}-coordinating histidine and the conserved phenilalanine in the middle of the α-helix. We have shown that a single point mutation of threonine 202 to leucine in TIS11d is sufficient to decrease the affinity for Zn^{2+} of the C-terminal zinc finger and destabilize the structure, while the corresponding mutation of leucine 152 to threonine in TTP is sufficient to increase zinc binding affinity of ZF2. Second, we mutated the conserved phenylalanine 200 to alanine in TIS11d, to verify that its stacking against the Zn^{2+}-coordinating histidine is essential for the folding of ZF2. Indeed, this mutant is partially unstructured and ZF2 is unable to fold even upon addition of RNA. These results unequivocally verify the critical residues and interactions identified using MD that are necessary to increase Zn^{2+}-binding affinity and stablize the fold of ZF2.

The sequence alignment of 14,851 CCCH-type zinc finger domains shows that the residue three positions away from the first cysteine (i.e. Cys^1+3) is likely to be aromatic (Phe, Trp, Tyr or His) with a probability $> 94\%$. In addition, experimental studies of the F200A mutant of TIS11d confirm that, in the absence of this aromatic residue, the C-terminal zinc finger is unable to stably coordinate Zn^{2+}. These results suggest that most CCCH-type zinc finger proteins employ $\pi - \pi$ interactions to stabilize the Zn^{2+}-coordinating histidine in a rotameric state that is compatible with the tetrahedral geometry of the Zn^{2+}-binding site. The strategy adopted to maintain the stacking of the aromatic side chain with the Zn^{2+}-coordinating histidine is not conserved, however. Previous studies had shown that the extent of disorder of the

RNA-binding domain affects the activity of the protein in the cell [176]. Of the 14,851 CCCH-type zinc finger sequences that we examined, roughly half of them support the formation of hydrogen bonds using the O_γ from residues located either at Cys^1+5 or Cys^1+6: the remaining zinc fingers may use alternative mechanisms to stabilize the coordination of Zn^{2+}. Through this apparent variety of Zn^{2+}-coordination stabilizing mechanisms, evolution can modulate the thermodynamic stability for this class of zinc fingers and, ultimately, regulate their biological activity.

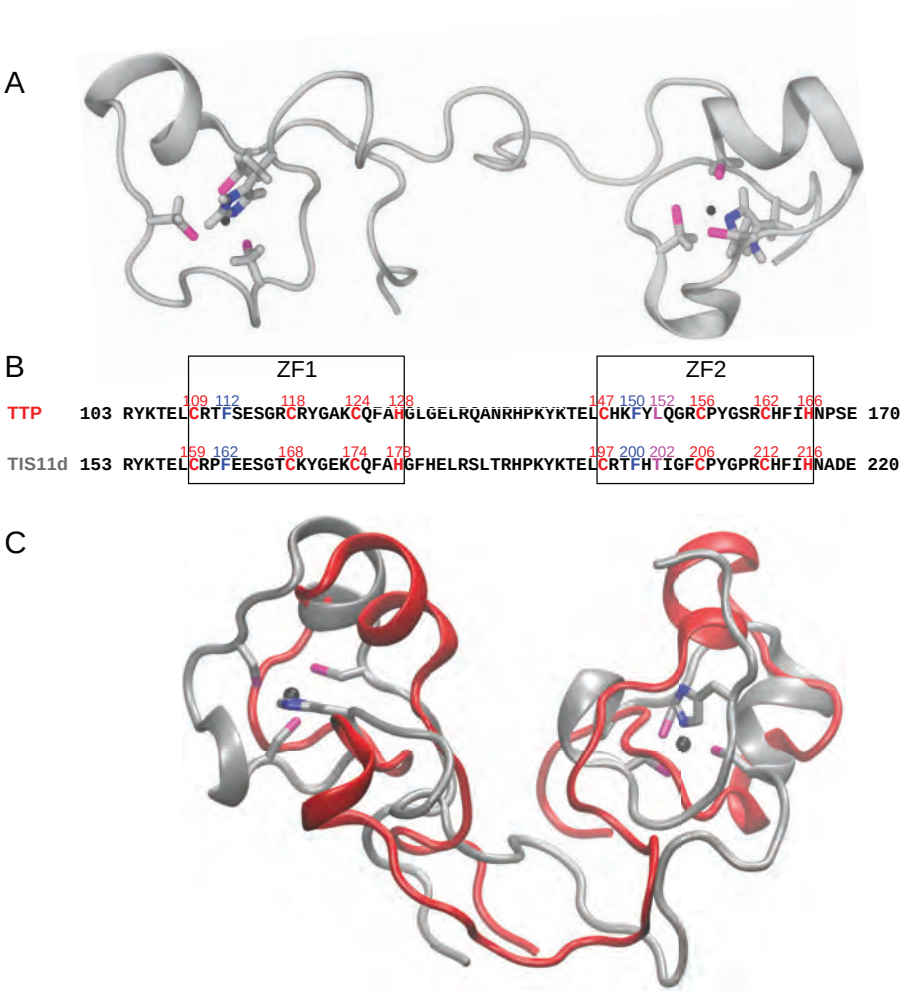


Figure 2.10: A: NMR solution structure of TIS11d (pdb entry: 1RGO). The side chains of the Zn^{2+}-coordinating residues are shown. Sulfur atoms are depicted in magenta, nitrogen atoms in blue and Zn^{2+} in black. B: Sequence alignment of the tandem zinc finger domains of TIS11d and TTP is shown. Zn^{2+}-coordinating residues are depicted in red, conserved phenylalanines at position Cys^1+3 are depicted in blue. Boxes indicate each zinc finger. C: Equilibrated structures of TIS11d (gray) and TTP (red) from MD trajectories. The side chains of the Zn^{2+}-coordinating residues of TIS11d are shown. Sulfur atoms are depicted in magenta, nitrogen atoms in blue and Zn^{2+} in black. Overall RMSD between the equilibrated structures of the two proteins is in average 8.8 Å, whereas ZF1 and ZF2 show a RMSD < 1 Å and < 1.5 Å, respectively.

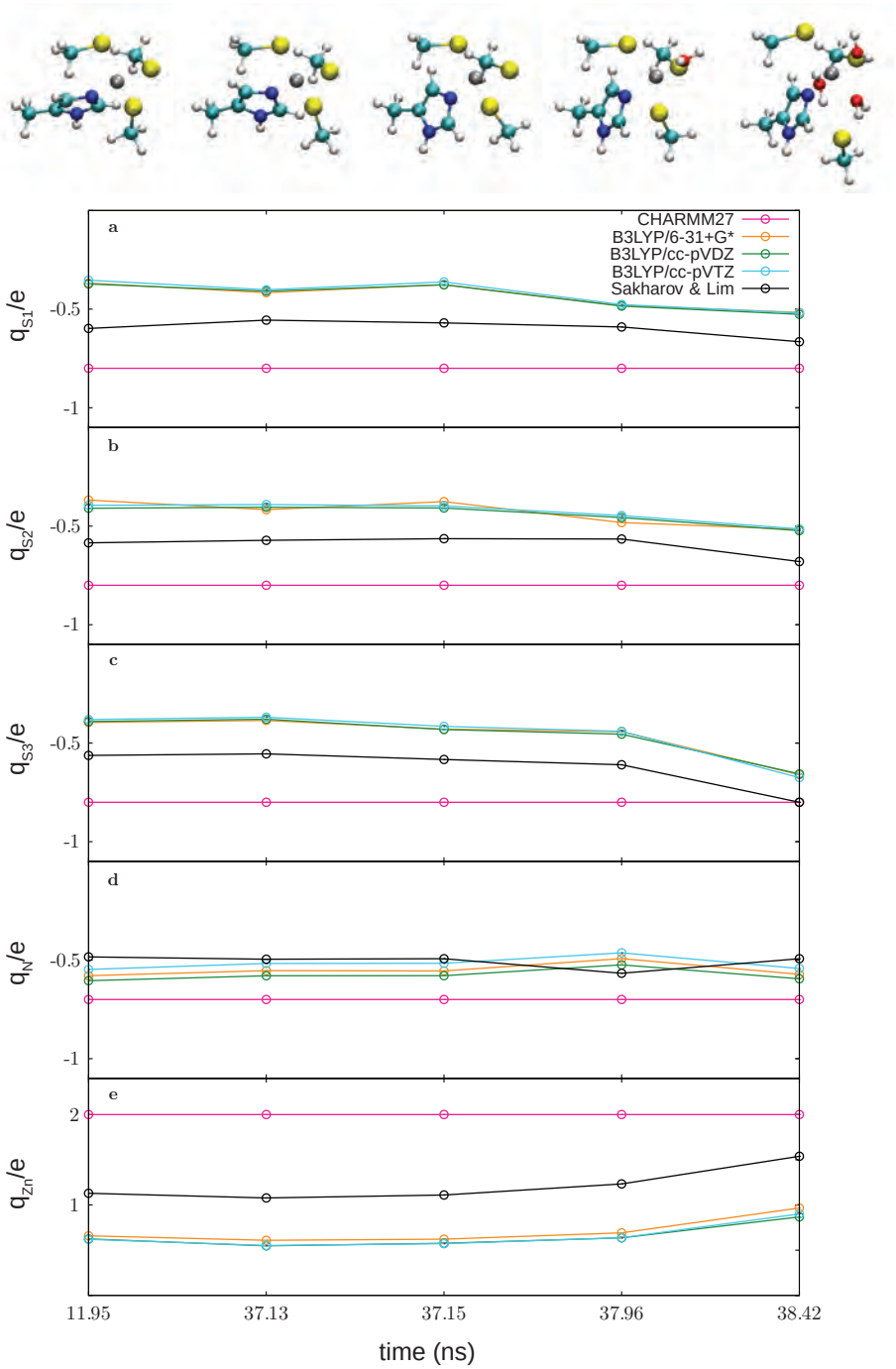

CHARMM27
B3LYP/6-31+G*
B3LYP/cc-pVDZ
B3LYP/cc-pVTZ
Sakharov & Lim
a
b
c
d
e
q_{S1}/e
q_{S2}/e
q_{S3}/e
q_N/e
q_{Zn}/e
-0.5
-1
2
1
11.95
37.13
37.15
37.96
38.42
time (ns)

Figure 2.11: Classical polarizable charge-transfer model approximates environmental dependence of electrostatic charges in the zinc fingers of TTP/TIS11d throughout molecular dynamics simulations. In order to evaluate the accuracy of the zinc finger model of Sakharov and Lim [181] in classical molecular dynamics simulations of TTP and TIS11d, five configurations of ZF2 in TTP were extracted from a single trajectory to represent different states of zinc coordination (these configurations are visualized at the top of the figure above), ranging from fully coordinated by the CCCH residues (leftmost three configurations) to partially coordinated (rightmost two configurations). As a reference, the charges were computed using density functional theory with the B3LYP approximate exchange-correlation functional [182, 183, 184] using Gaussian09 [185], after truncating the protein backbone atoms from the cysteine and histidine residues to form methanethiolate anions and a 4-methylimidazole molecule, respectively. The charges from the CHARMM27 [186] force field are compared with those from the polarizable charge-transfer model of Sakharov and Lim [181] along with the natural bond orbital (NBO) charges [187, 188, 189, 190] corresponding to a series of different basis sets. The 6-31+G* basis set was used in the parameterization of the CHARMM27 [186] force field, along with that of the Sakharov and Lim [181] model. Shown for comparison are the cc-pVDZ and cc-pVTZ basis sets. Agreement among the different basis sets suggests that these choices are adequate, as basis set expansion yields very small changes in the charges. Panels (a-c) report the charges on the sulfur atom of the methanethiolate anion, panel (d) reports the charge on the zinc-coordinating nitrogen atom of the 4-methylimidazole molecule, while (e) reports the charge on the zinc cation. In each configuration, the polarizable charge-transfer model represents an improved approximation to the fixed-charge model.

	RMSD(protein) (Å)	RMSD(ZF1) (Å)	RMSD(ZF2) (Å)
TIS11d r1	7.4 ± 2.3	0.8 ± 0.1	0.6 ± 0.1
TIS11d r2	7.1 ± 2.2	0.7 ± 0.2	0.7 ± 0.2
TIS11d r3	3.3 ± 0.6	0.5 ± 0.1	0.5 ± 0.1
TIS11d r4	3.8 ± 1.7	0.5 ± 0.1	0.6 ± 0.2
TIS11d r5	5.4 ± 2.5	0.6 ± 0.1	0.6 ± 0.1
TIS11d r6	1.6 ± 0.7	0.6 ± 0.1	0.5 ± 0.1
TTP r1	5.2 ± 1.2	0.5 ± 0.1	1.4 ± 0.4
TTP r2	7.5 ± 1.8	0.6 ± 0.1	0.7 ± 0.1
TTP r3	4.3 ± 2.7	0.6 ± 0.2	0.6 ± 0.2

Table 2.1: Values of RMSD for the equilibrated structures of the protein and ZF1 and ZF2, calculated for six trajectories 100 ns long of TIS11d and three trajectories 100 ns long for TTP. Values are shown as mean ± STD.

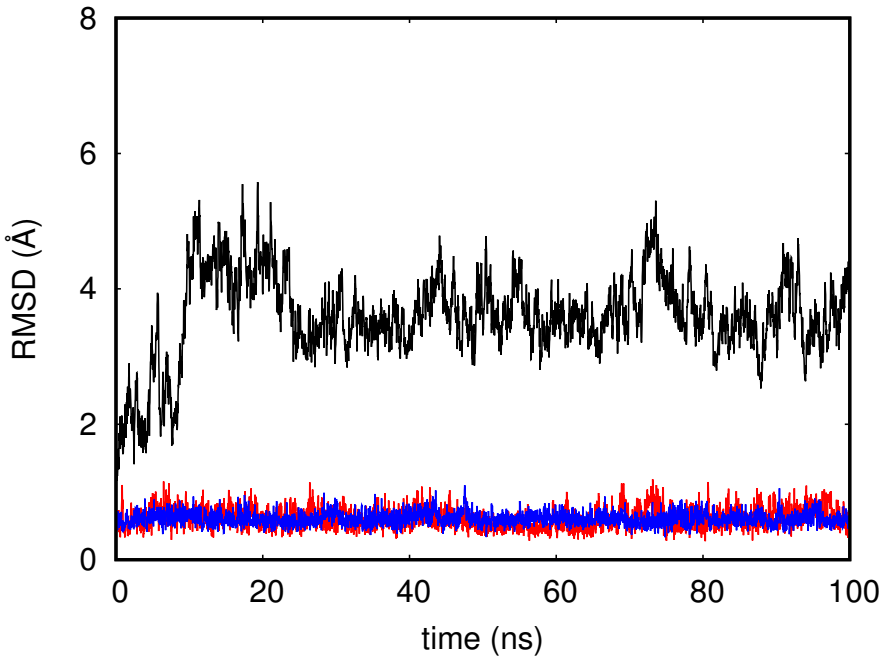


Figure 2.12: RMSD as a function of time for a representative 100 ns long trajectory of TIS11d. RMSD for the overall protein is depicted in black, for ZF1 in red and for ZF2 in blue.

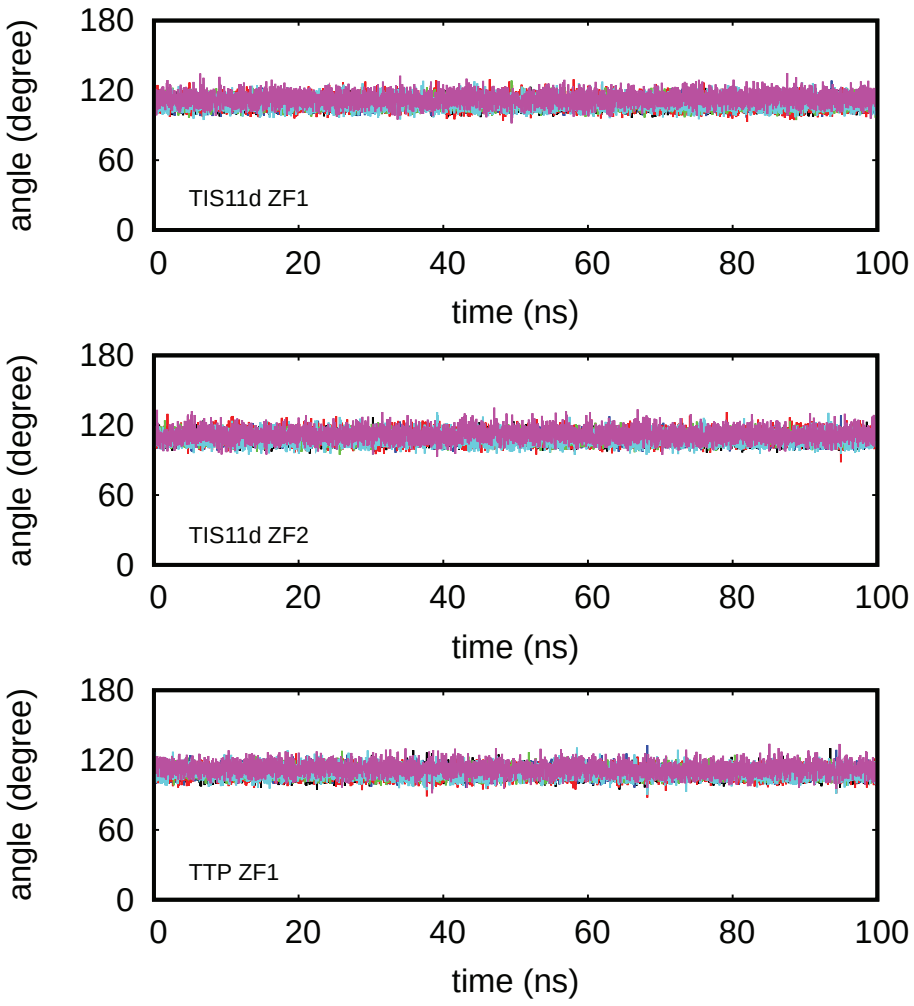


Figure 2.13: The geometry of the zinc coordination in both zinc fingers of TIS11d and in ZF1 of TTP is monitored. The angles between the zinc ion and the zinc coordinating atoms are shown for a representative 100 ns long trajectory of TIS11d and for the 100 ns long trajectory of TTP where zinc coordination is lost: $\measuredangle \mathrm{S}^{Cys^1}$-$\mathrm{Zn}^{2+}$-$\mathrm{S}^{Cys^2}$ in black, $\measuredangle \mathrm{S}^{Cys^1}$-$\mathrm{Zn}^{2+}$-$\mathrm{S}^{Cys^3}$ in blue, $\measuredangle \mathrm{S}^{Cys^1}$-$\mathrm{Zn}^{2+}$-$\mathrm{N}_{\varepsilon}^{His^4}$ in red, $\measuredangle \mathrm{S}^{Cys^2}$-$\mathrm{Zn}^{2+}$-$\mathrm{S}^{Cys^3}$ in green, $\measuredangle \mathrm{S}^{Cys^2}$-$\mathrm{Zn}^{2+}$-$\mathrm{N}_{\varepsilon}^{His^4}$ in cyan and $\measuredangle \mathrm{S}^{Cys^3}$-$\mathrm{Zn}^{2+}$-$\mathrm{N}_{\varepsilon}^{His^4}$ in magenta.

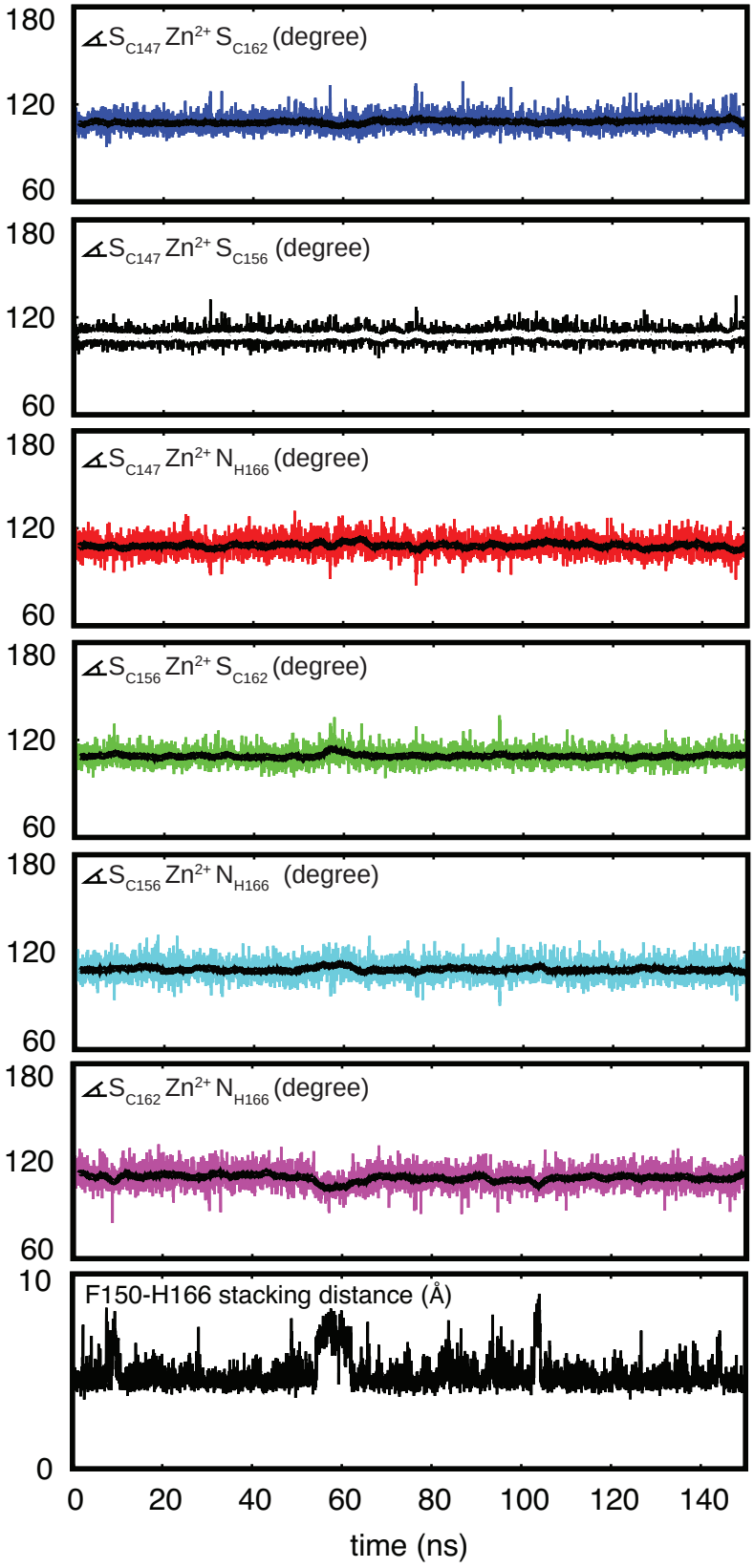

180
120
60
∡S$_{C147}$ Zn^{2+} S$_{C162}$ (degree)
∡S$_{C147}$ Zn^{2+} S$_{C156}$ (degree)
∡S$_{C147}$ Zn^{2+} N$_{H166}$ (degree)
∡S$_{C156}$ Zn^{2+} S$_{C162}$ (degree)
∡S$_{C156}$ Zn^{2+} N$_{H166}$ (degree)
∡S$_{C162}$ Zn^{2+} N$_{H166}$ (degree)
10
0
F150-H166 stacking distance (Å)
0
20
40
60
80
100
120
140
time (ns)

Figure 2.14: The geometry of the zinc coordination in the C-terminal zinc finger of TTP is monitored. The angles between the zinc ion and the zinc coordinating atoms are shown for the first trajectory of TTP, 150 ns long: $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C156}$ in black, $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C162}$ in blue, $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{N}_{\varepsilon}^{H166}$ in red, $\measuredangle \mathrm{S}^{C156}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C162}$ in green, $\measuredangle \mathrm{S}^{C156}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{N}_{\varepsilon}^{H166}$ in cyan and $\measuredangle \mathrm{S}^{C162}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{N}_{\varepsilon}^{H166}$ in magenta. Stacking distance between F150 and H166 is shown as a function of time. The distance between the aromatic rings was calculated as the distance between the centers of mass for the heavy atoms of the two side chains. The running average of the angles, calculated for a window of 100 datapoints, is shown as a black line (white for $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C156}$) within each plot.

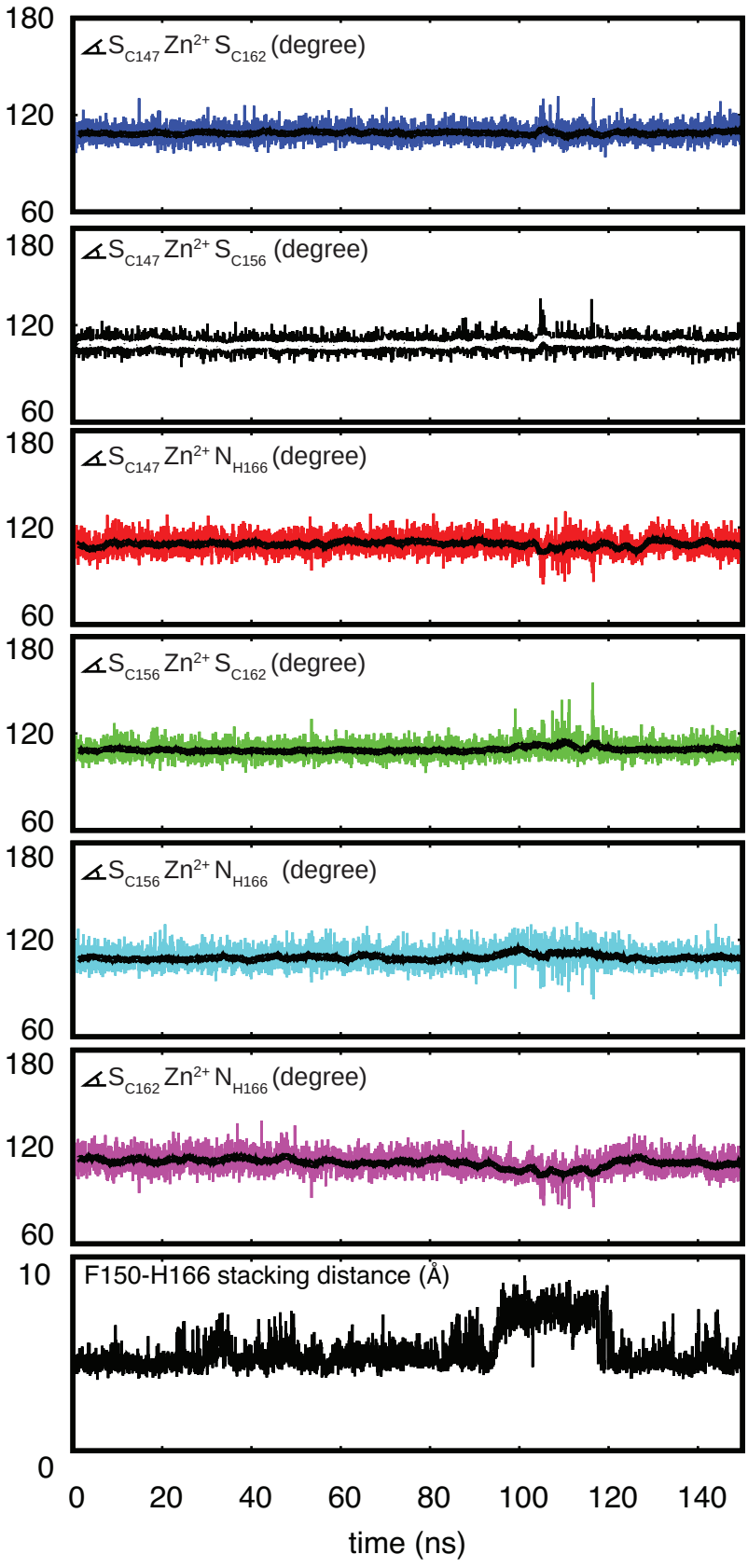

180
120
60
∡S$_{C147}$ Zn^{2+} S$_{C162}$ (degree)
∡S$_{C147}$ Zn^{2+} S$_{C156}$ (degree)
∡S$_{C147}$ Zn^{2+} N$_{H166}$ (degree)
∡S$_{C156}$ Zn^{2+} S$_{C162}$ (degree)
∡S$_{C156}$ Zn^{2+} N$_{H166}$ (degree)
∡S$_{C162}$ Zn^{2+} N$_{H166}$ (degree)
10
0
F150-H166 stacking distance (Å)
0
20
40
60
80
100
120
140
time (ns)

Figure 2.15: The geometry of the zinc coordination in the C-terminal zinc finger of TTP is monitored. The angles between the zinc ion and the zinc coordinating atoms are shown for the second trajectory of TTP, 150 ns long: $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C156}$ in black, $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C162}$ in blue, $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{N}_{\varepsilon}^{H166}$ in red, $\measuredangle \mathrm{S}^{C156}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C162}$ in green, $\measuredangle \mathrm{S}^{C156}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{N}_{\varepsilon}^{H166}$ in cyan and $\measuredangle \mathrm{S}^{C162}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{N}_{\varepsilon}^{H166}$ in magenta. Stacking distance between F150 and H166 is shown as a function of time. The distance between the aromatic rings was calculated as the distance between the centers of mass for the heavy atoms of the two side chains. The running average of the angles, calculated for a window of 100 datapoints, is shown as a black line (white for $\measuredangle \mathrm{S}^{C147}\text{-}\mathrm{Zn}^{2+}\text{-}\mathrm{S}^{C156}$) within each plot.

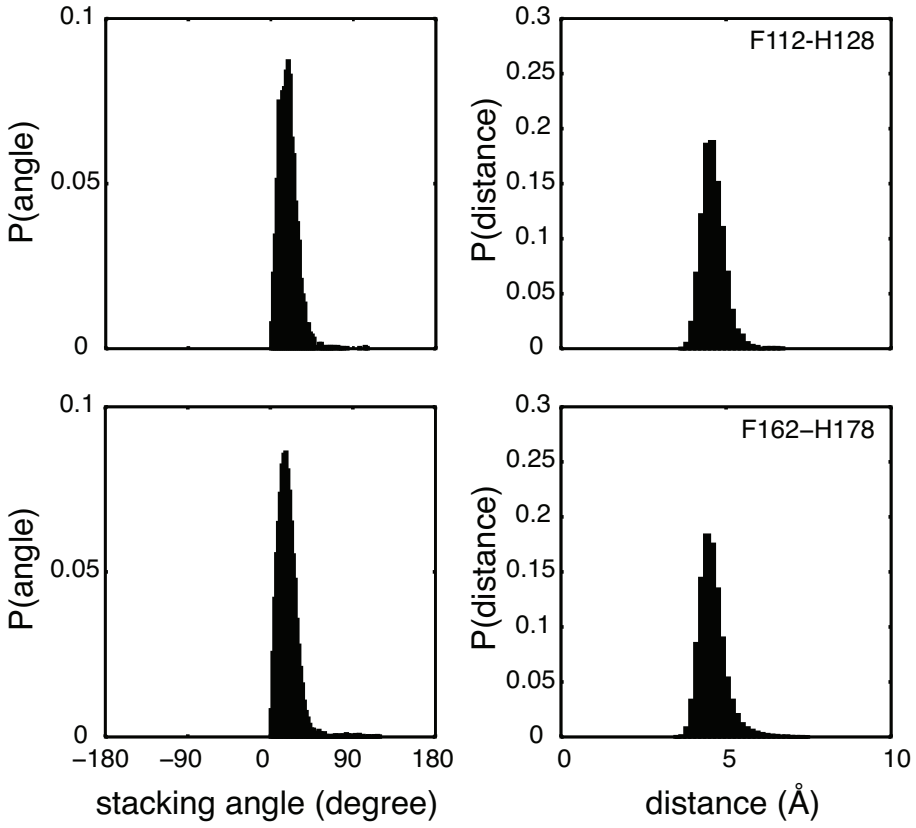


Figure 2.16: Aromatic side chain stacking of phenylalanine and histidine in ZF1 of TTP (top) and TIS11d (bottom). Probability distribution of the stacking angle (left), calculated as the angle between the normals of the two aromatic rings (the planes for the side chains are defined by atoms C_{δ_2}, C_{ε_1} and N_{ε_2} for histidine and C_ζ, C_{ε_2},C_{δ_2} for phenylalanine). Probability distribution of the distance between the aromatic rings (right), calculated as the distance between the centers of mass of the heavy atoms of the two side chains. Data refer to the 100 ns long unfolding MD trajectory for TTP and to six 100 ns long MD trajectory for TIS11d.

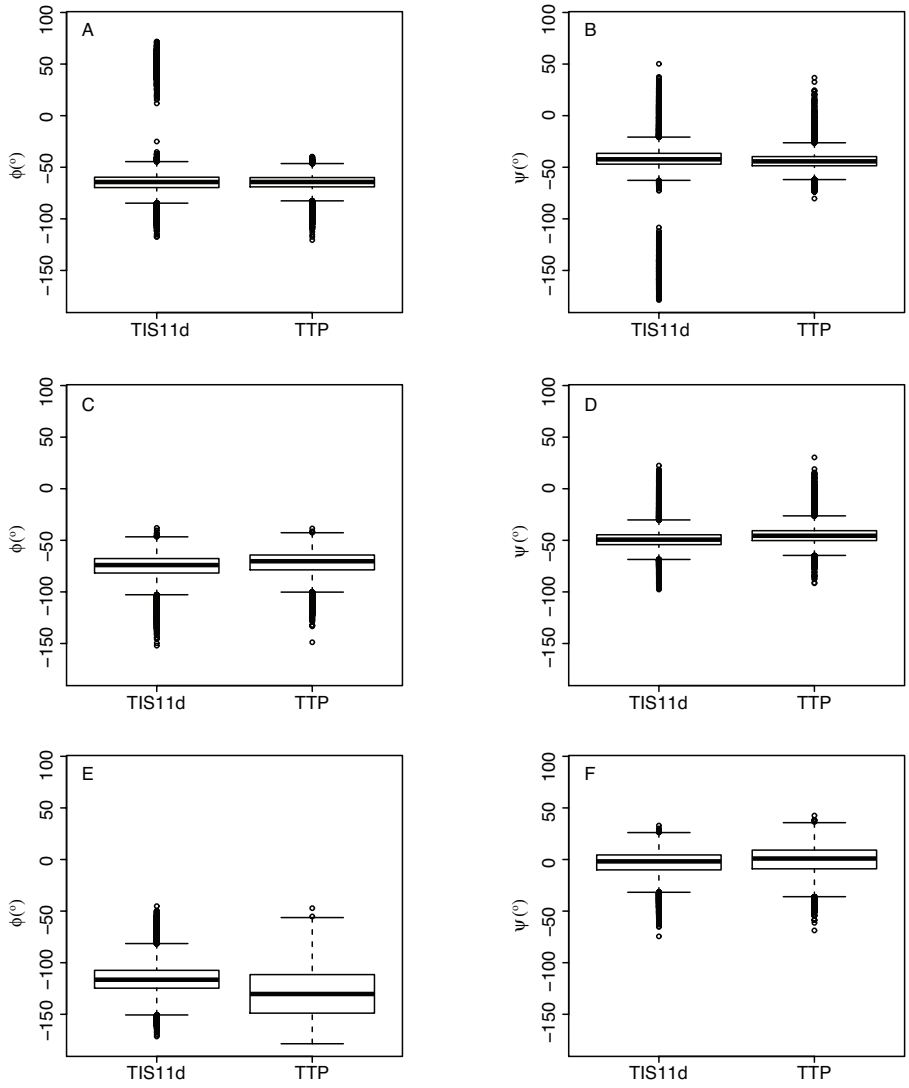


Figure 2.17: Box-and-whisker plots of the distribution of ϕ and ψ angles for residues 201 (A,B), 202 (C,D), 203 (E,F) of TIS11d and 151 (A,B), 152 (C,D), 153 (E,F) of TTP. Data refer to the 100 ns long unfolding MD trajectory for TTP and to six 100 ns MD trajectories for TIS11d. The outlier points in panels A and B for TIS11d, centered at $\phi = 50°$ and $\psi = -150°$, correspond to a single event observed in one trajectory where a hydrogen bond within the α-helix is transiently lost.

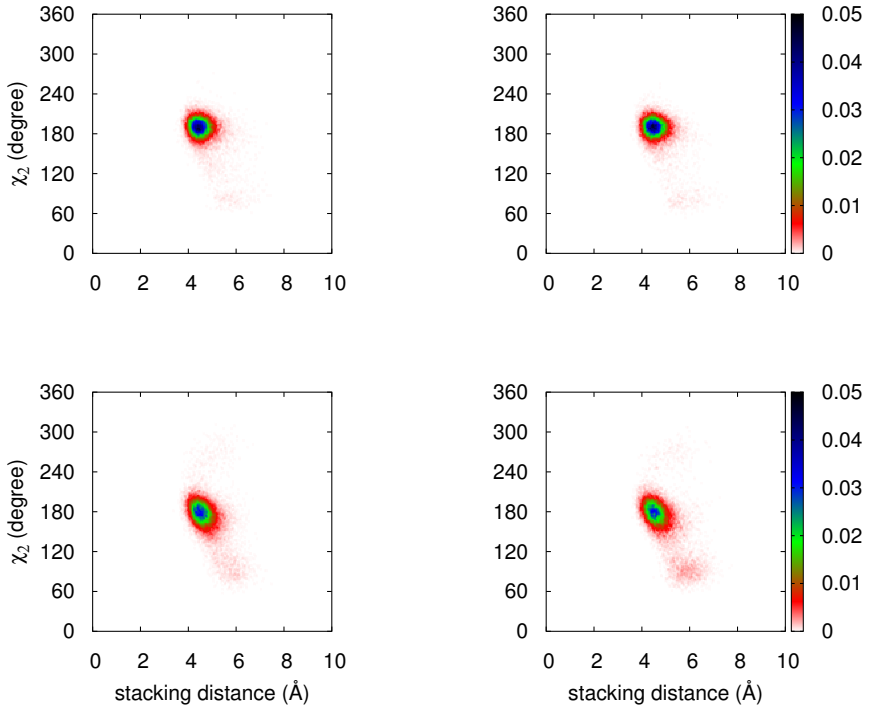


Figure 2.18: The T202L mutant form of TIS11d samples the Zn^{2+} binding-incompatible state with higher probability than wild type. Probability density distribution of the dihedral angles χ_2 of H178 and the stacking distance of F162 and H178 in ZF1 of TIS11d (top left), χ_2 of H216 and the stacking distance of F200 and H216 in ZF2 of TIS11d (bottom left), χ_2 of H128 and the stacking distance of F112 and H128 in ZF1 of TIS11d T202L (top right), χ_2 of H166 and the stacking distance of F150 and H166 in ZF2 of TIS11d T202L (bottom right). Data are taken from six 100 ns MD trajectories of TIS11d wild type and from six 100 ns MD trajectories of TIS11d T202L. The color bars show the values of the probability density calculated for χ_2 and stacking distance as the number of counts normalized by the total number of observations and by the area of each bin.

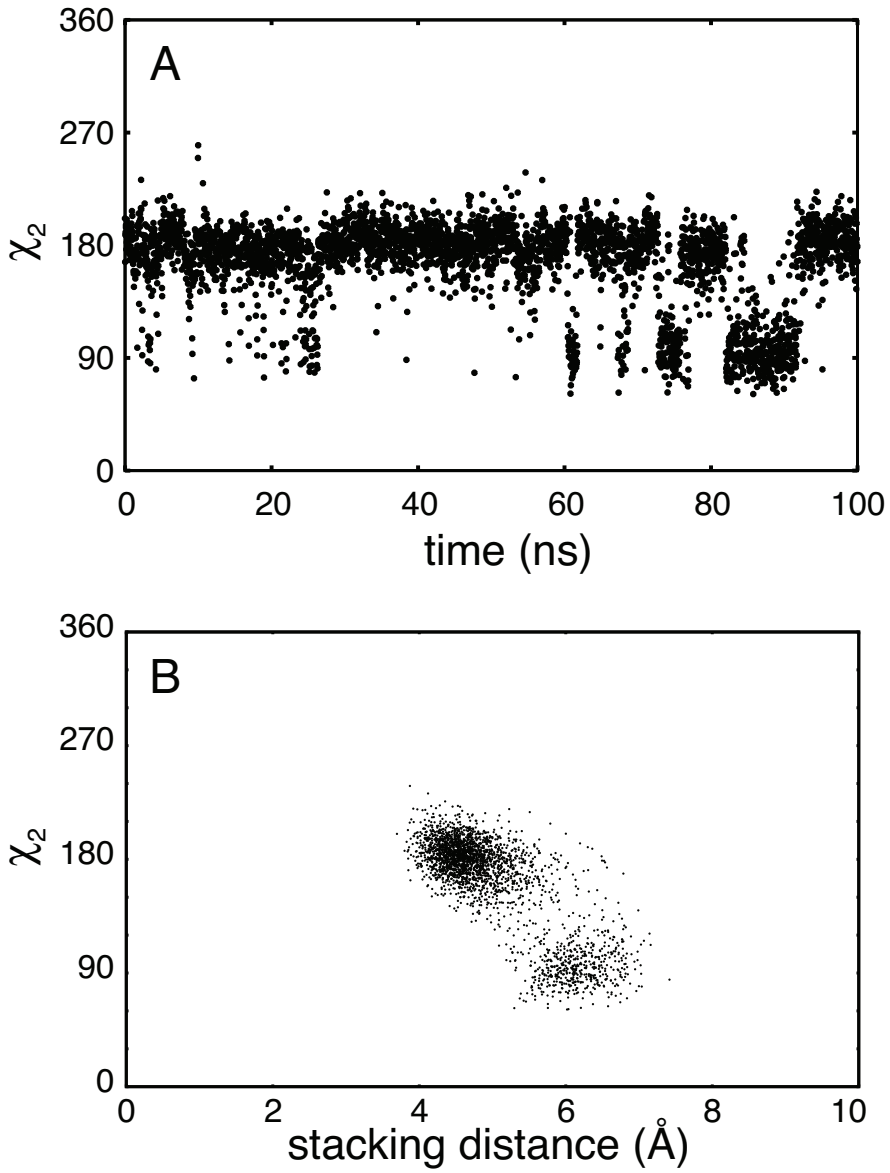


Figure 2.19: A: The χ_2 dihedral angle of the side chain of H216 is shown as a function of time for a representative 100 ns long trajectory of the TIS11d T202L mutant. B: Scatter plot of the stacking distance between the aromatic ring of F200 and H216 and the χ_2 dihedral angle of H216 for the TIS11d T202L mutant. The distance between the aromatic rings was calculated as the distance between the centers of mass for the heavy atoms of the two side chains. Data were extracted from a representative 100 ns long trajectory of the TIS11d T202L mutant.

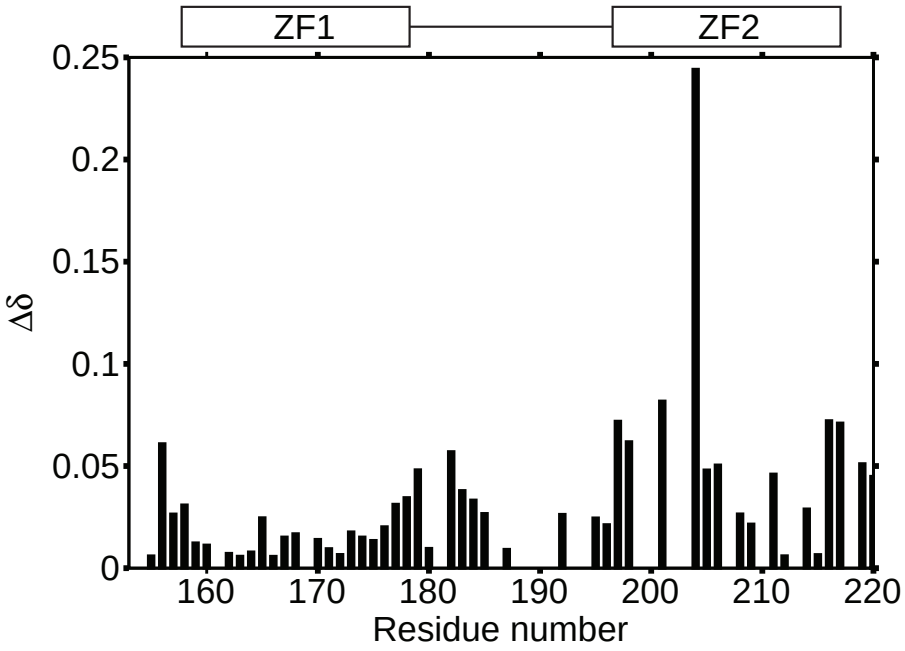


Figure 2.20: Chemical shift difference between the ^{15}N-^{1}H HSQC spectra of TIS11d T202L and TIS11d wild type. $\Delta\delta$ was calculated as $\sqrt{\left(\Delta\delta_{^{15}N} \cdot \frac{\gamma_{^{15}N}}{\gamma_{^{1}H}}\right)^2 + \left(\Delta\delta_{^{1}H}\right)^2}$.

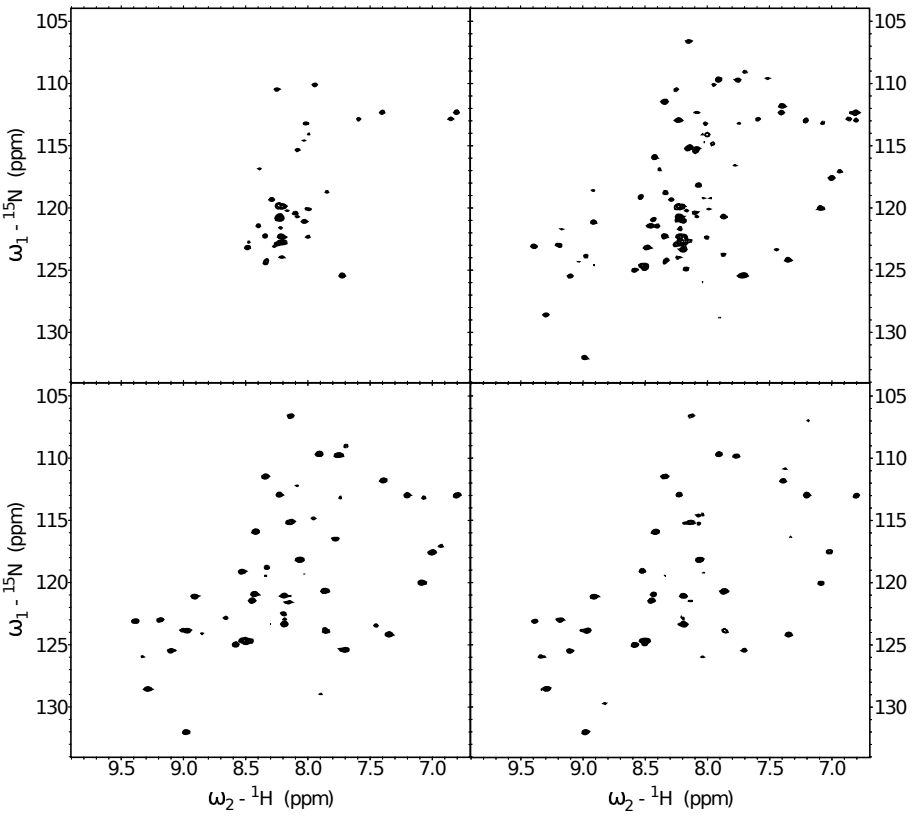


Figure 2.21: Zn^{2+} titration of TIS11d T202L followed by NMR spectroscopy. The ^{15}N-^{1}H HSQC spectra correspond to the protein in absence of zinc (top left), protein with 0.5 equivalents of zinc in solution (top right), protein with 1 equivalent of zinc in solution (bottom left) and protein with 3-fold excess of zinc (bottom right). Cross-peaks from ZF2 have low intensities, close to the noise.

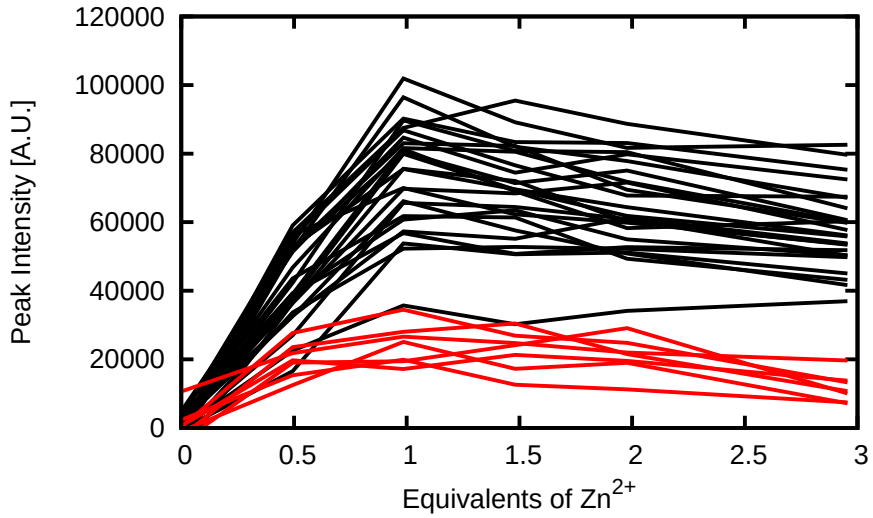


Figure 2.22: Zn^{2+} titration of TIS11d T202L followed by NMR spectroscopy. Cross-peak intensities from the ^{15}N-^{1}H HSQC spectra are shown as a function of zinc concentration. Cross-peak intensities of residues 155-186 are depicted in black, 187-220 in red. The overall decrease in intensity observed at high zinc concentration is due to a small change in pH upon $ZnSO_4$ addiction.

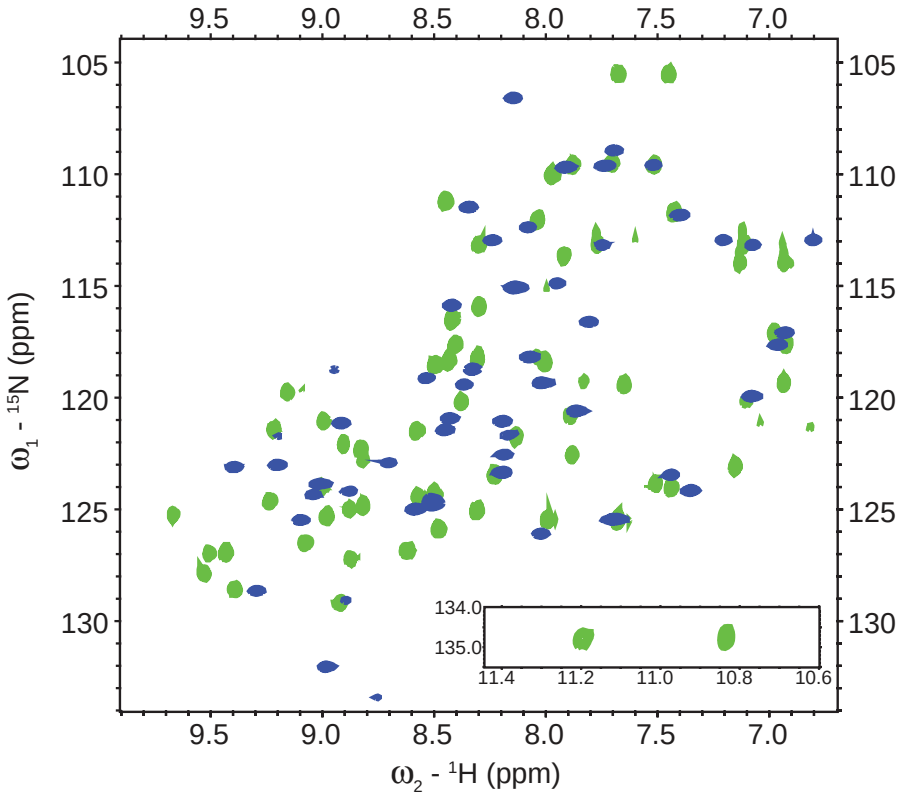


Figure 2.23: ^{15}N-^{1}H HSQC spectra of TIS11d T202L free (blue) and bound (green) to 5′-UUUAUUUAUUUU-3′ RNA.

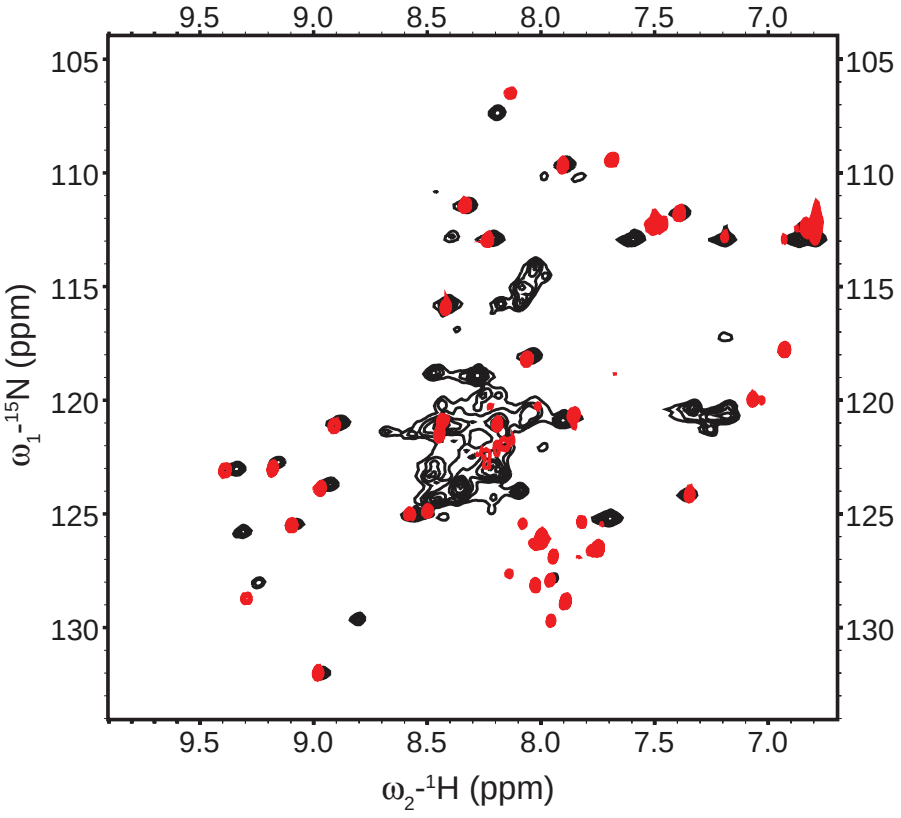


Figure 2.24: Comparison of ^{15}N-^{1}H HSQC spectra of TIS11d F200A (red) and TIS11d C212S (black).

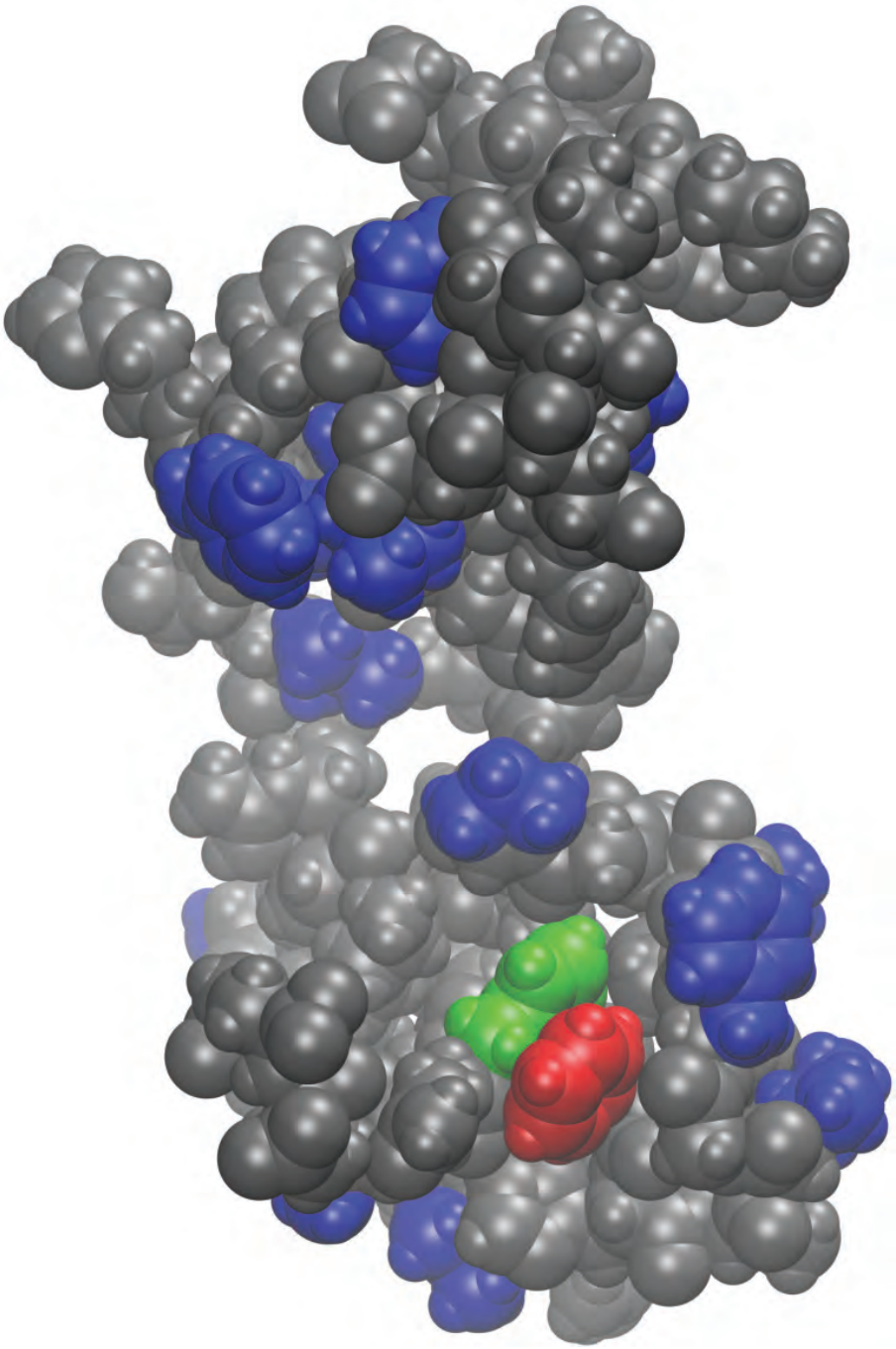


Figure 2.25: All atoms representation of the NMR solution structure of TIS11d (pdb entry: 1RGO). Side chains of hydrophobic residues are depicted in blue, side chain of F200 in red and side chain of H216 in green.

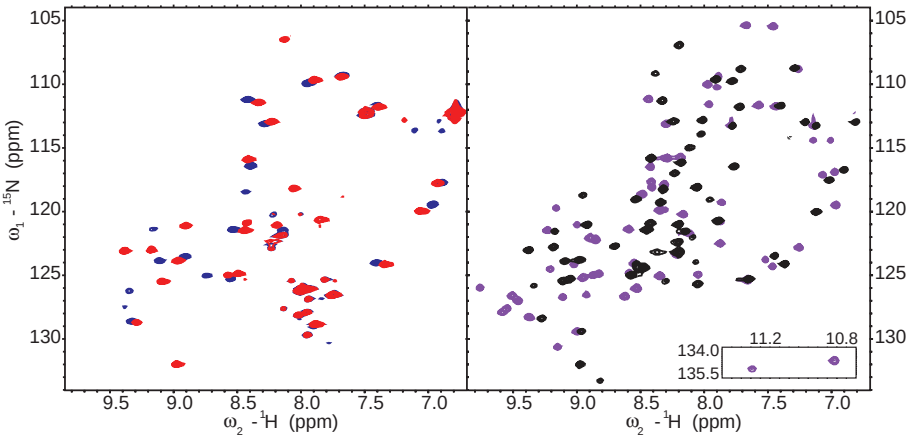


Figure 2.26: ^{15}N-^{1}H HSQC spectra of TIS11d F200A free (red) and bound (blue) to 5′-UUUAUUUAUUUU-3′ RNA compared to the spectra of TIS11d wild type free (black) and bound (purple) to 5′-UUUAUUUAUUUU-3′ RNA.

2.5 Material and Methods

2.5.1 TTP and TIS11d RNA-binding domain homology model building and preparation

The unknown structure of the ligand-free RNA-binding domain of TTP (residues 101 to 170) was generated starting from the lowest energy NMR structure of TIS11d (pdb entry: 1RGO) bound to ARE (5′-UUAUUUAUU-3′) using the SWISS-MODEL Server [191, 192, 193]. The resulting structure was solvated using VMD 1.9.2 [194] in an orthorhombic water box (50x66x52 Å). Six Cl^- ions were added to the system to neutralize the charge. Similarly, to obtain the structure of TIS11d in the apo state, the lowest energy NMR structure of TIS11d (residues 151 to 220) was solvated, after removal of the RNA molecule, in a orthorhombic water box (60x75x66 Å) containing a single Cl^- ion to enforce charge neutrality. The size of the water box used to solvate TIS11d needed to be larger than that used for TTP. This was to avoid the interaction of the protein with its image in the surrounding boxes due to an infrequently observed extention of the linker region. This rare linker extension was never observed for TTP. Two additional simulations of 30 ns were collected for TTP solvated in the same size water box used for TIS11d (60x75x66 Å). As expected, the size of the box did not affect the structure and dynamics of TTP in solution. The mutation of threonine 202 into leucine (T202L) was introduced in the sequence of wild type TIS11d using the Mutator plugin (v. 1.3) of VMD [194].

2.5.2 Simulation Protocol

The solvated proteins, described above, were energy-minimized and equilibrated using the NAMD 2.10 molecular modeling package [195] and the CHARMM27 force field [186]. The force field was modified following Sakharov and Lim [181] in order to include polarization and charge transfer effects for the Zn^{2+} ions and the side chain atoms of the zinc coordinating residues. Simulations including a fixed charge

non-bonded model of Zn^{2+} were not able to accurately reproduce the tetrahedral coordination geometry for the zinc ions in TIS11d [172]. In order to further assess the accuracy of the Sakharov and Lim polarizable charge transfer model for these CCCH-family tandem zinc fingers, five molecular configurations of ZF2 from TTP, representing diverse coordination states for the Zn^{2+} cation, were selected and examined using electronic structure calculations and a series of levels of theory. The charges transferred in the model of Sakharov and Lim reproduce accurate quantum chemical charges reasonably well and substantially improve upon the accuracy of the fixed charges from CHARMM27 [186] (see Fig. 2.11). Prior to equilibration, all systems were subjected to energy minimization in three stages, with restraints sequentially removed: first, all heavy atoms were constrained; next, only C_α atoms were constrained and finally, minimization was done without constraints. The systems were subsequently subjected to step-wise heating during constant volume MD with restraints applied to C_α atoms, followed by 10 ps of unconstrained constant-NPT molecular dynamics equilibration at 1 atm and 298 K. Trajectories were subsequently collected from constant-NPT MD simulations at 1 atm and 298 K. Temperature and pressure were maintained using Langevin dynamics (damping coefficient: 5 ps^{-1}) and the Nosé-Hoover Langevin piston method, respectively. The equations of motion were integrated using the SHAKE constraint algorithm in order to use a 2 fs time step [196]. Non-bonded interactions were calculated at every time step with a cut-off distance of 12 Å and a switching distance of 10 Å. The particle mesh Ewald method was used to treat electrostatic interactions with periodic boundary conditions [197, 198]. Three trajectories of TTP were run, each for a total of 100 ns, with the last 80 ns used for data collection and analysis. Two of the three trajectories of TTP, where the Zn^{2+} ion remains bound to ZF2, were extended for a further 50 ns, to a total of 150 ns. Loss of Zn^{2+} from ZF2 was not observed in the extended trajectories. Six trajectories of TIS11d were run, each for a total of 100 ns, with the last 80 ns used for data collection

and analysis. Six trajectories were run for the T202L mutant of TIS11d, each for a total of 100 ns, with the last 80 ns considered for data collection and analysis. Although the thermodynamic values for each trajectory of TTP and TIS11d were equilibrated in the initial configurations, on average it took about 20 ns to reach structural equilibration. Conformations were judged equilibrated as quantified by root mean square displacement (RMSD) from the original structure. The structure of the zinc fingers is well maintained for the entire duration of the simulations, while the linker region remains flexible, as shown by the average RMSD calculated for the full protein and the zinc fingers, see Figs. 2.12, 2.13, 2.14, 2.15 and Table 2.1. Trajectories were analyzed using VMD 1.9.2 [194] and molecular configurations were visualized using STRIDE [199] and Tachyon [200].

2.5.3 Protein Expression

The RNA-binding domain of human TIS11d (residues 152-220) and TTP (residues 102 to 170) were synthesized by Genescript and cloned into a modified pet28 vector with a SUMO tag between BamH1 restriction site. The F200A, T202L and C212S mutations of TIS11d and the L152T mutation of TTP were generated via Quikchange mutagenesis. TIS11d wild type, F200A, T202L and C212S mutants and TTP wild type and L152T mutant were expressed within BL21(DE3) *E. coli* competent cells. Isotopic labeling with ^{15}N was performed by growing the cells in M9 containing 1 g of $^{15}NH_4Cl$ per liter. The cells were grown at 37°C to an OD_{600} of 0.8 and then induced for 4 hours with 1 mM Isopropyl β-D-1 thiogalactopyranoside (IPTG) and 0.1 mM $ZnSO_4$ at the same temperature. Harvested cells were lysed using a cell disruptor in 50 mL buffer containing 50mM Tris HCl, pH 8.0, 50 mM NaCl, and 1 EDTA free cOmplete^{TM} protease inhibitor tablet (Roche). Lysates were centrifuged at 19500 RPM for 1 hour at 4°C and passed through a 20 mL pre-packed PrepEase His Tagged resin (Affymetrix), washed with 5 column volumes of 50 mM Tris HCl,

pH 8.0, 50 mM NaCl, 5 mM imidazole, and eluted with 50 mM Tris HCl, pH 8.0, 50 mM NaCl, 350 mM imidazole. The SUMO tag was cleaved off with ULP1. The cleavage reaction was performed for 2 hours at room temperature, using a ULP1-to-protein ratio of 1:10. The protein was then passed through a 5 mL HiTRAP Q and SP column (GE Healthcare Life Sciences) pre-equilibrated with a buffer containing 50 mM Tris HCl, pH 8.0, 50 mM NaCl. Purified protein solution was buffer exchanged into 10 mM Tris, pH 6.2, 20 mM KCl, 2 mM DTT, 0.1 mM $ZnSO_4$ by dialysis and concentrated using a 3 KDa Centriprep concentrator (Millipore).

2.5.4 CD spectroscopy

Far-UV circular dichroism (CD) spectra were recorded for TIS11d wild type, F200A and T202L mutants and for TTP wild type and L152T mutant in 50 mM HEPES, pH 7.0, 20 mM KCl and 1 mM TCEP using a Jasco-810 spectropolarimeter (Jasco Inc., Easton, MD). Curves were monitored from 200-260 nm in a 0.1 cm path length quartz cuvette using a scan rate of 20 nm min^{-1} and a response time of 8 s. The sample temperature for all CD measurements was maintained at 293 K.

2.5.5 NMR spectroscopy

Folding of TIS11d wild type, F200A and T202L mutants and TTP wild type and L152T mutant was monitored via NMR spectroscopy. ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) spectra were collected at 298 K on a Varian Inova spectrometer operating at 600 MHz equipped with a triple-resonance cold probe. Data processing was performed using NMRPipe [201] and Sparky [202] software.

CHAPTER 3

A DISORDER-TO-ORDER TRANSITION REGULATES THE FUNCTION OF THE *CAENORHABDITIS ELEGANS* RNA-BINDING PROTEIN MEX-5

3.1 Abstract

CCCH-type tandem zinc finger (TZF) domains are found in many RNA-binding proteins (RBPs) that regulate the essential processes of post-transcriptional gene expression and splicing through direct protein-RNA interactions. In *Caenorhabditis elegans*, RBPs control the translation, stability, or localization of maternal mRNAs required for patterning decisions prior to zygotic gene activation. Here, we report that the CCCH-type TZF domain of MEX-5, a *C. elegans* protein that leads a cascade of RBP localization events following fertilization, contains an unstructured zinc finger that folds upon binding of its RNA target. To evaluate if this disorder-to-order transition upon RNA-binding contributes to MEX-5 function, we designed a variant MEX-5 where both fingers are fully folded in the absence of RNA. We characterized the RNA-binding activity of this variant MEX-5 and we found that the binding affinity and specificity are unchanged compared to the wild type protein. We used CRISPR-hr to introduce this variant into the endogenous *mex-5* locus. Homozygotes are sterile, form massive uterine tumors within a few days of reaching adulthood, and often die by bursting. Heterozygotes are fertile but form tumors at advanced age. We observed that the tumors are derived from embryonic cells wherein nuclei divide, but not the cytoplasm, leading to giant polynucleated embryoid cells in the uterus. Together, our results show that the unfolded state of MEX-5 is critical to its function *in vivo* by a mechanism distinct from its RNA-binding activity.

3.2 Introduction

During embryogenesis a fertilized oocyte develops from a single cell, the zygote, into a multicellular organism. Before zygotic transcription begins, restricted distribution of maternal factors and their regulation coordinate early developmental processes [203, 204]. Maternal factors are maternally provided quiescent mRNAs and proteins present in the oocyte cytoplasm.

The *Caenorhabditis elegans* zygote is a classic model for the study of intracellular asymmetries [106, 205] in which conserved mechanisms of cell polarization and cytoplasmic organization have been elucidated [104]. After fertilization, polarization of the *C. elegans* zygote along the anterior-posterior (A-P) embryonic axis occurs through the action of several conserved polarity regulators, the maternal factors PAR proteins, which localize to the anterior (PAR-6, PAR-3, PKC-3) or posterior (PAR-2, PAR-1) cell cortex [206]. As a response to this asymmetry, cell-fate determinants become asymmetrically distributed in the cytoplasm. Among them, the CCCH-type tandem zinc finger (TZF) RNA-binding protein MEX-5 redistributes across the length of the 50 μm zygote with a concentration gradient anterior-high to posterior-low in a timespan of $\sim$ 10 min [108, 207]. Such anterior enrichment of MEX-5 is driven by a phosphogradient [208, 209, 111]: phosphorylation of MEX-5 by the kinase PAR-1 results in increased MEX-5 diffusion in the posterior, whereas in the anterior the phosphatase PP2A reverts MEX-5 to a slow-diffusing state, which formation requires MEX-5 RNA-binding [111]. The spatial asymmetry in the generation of the phosphorylated forms of MEX-5 generates a diffusion gradient that causes the protein to concentrate in the region of low diffusivity [104]. Subsequently, MEX-5 partitions other proteins such as PIE-1 and POS-1 to the posterior cytoplasm and PLK-1 to the anterior cytoplasm [210, 211, 212, 108]. The first stages of *C. elegans* development feature a series of rapid and invariant asymmetric cell divisions that determine the identity of the early embryonic blastomeres [104]. As a consequence of proper A-P

polarity, the zygote divides unequally, yielding the larger anterior cell AB and the smaller posterior cell P_1. Consequently, during the first cell division, the two daughter blastomeres inherit different determinants, which are involved in specifying their distinct fates (anterior/somatic and posterior/germline).

MEX-5 has been first identified by Schubert et al. in a genetic screen for maternal effect lethal mutants [108]. Schubert et al. observed that adults homozygous for null mutations in the *mex-5* gene produced embryos containing abnormally large numbers of muscles toward their anterior poles (Muscle EXcess, MEX). These embryos were unable to undergo body morphogenesis and died without hatching. In 2007, Pagano et al. characterized MEX-5 RNA-binding specificity and showed that the protein binds to poly-U streches, that are abundant in *C. elegans* $3'$ untranslated regions (UTRs) [110]. According to these results, it has been proposed [110] that MEX-5 may function, in the two-cell stage and during further cell division, as a broad-spectrum RNA-binding protein to activate maternal mRNA turnover in the AB blastomere and in the somatic lineage cells [213]. More recently, Han et al. have proposed a mechanism to explain the repulsive coupling between MEX-5 and POS-1 in the zygote, which is similar in many respects to that for MEX-5 segregation: the Polo-like Kinase PLK-1 interacts with MEX-5, distributed along an anterior-rich cytoplasmic gradient, and is recruited into slow-diffusing complexes; in the anterior cytoplasm PLK-1, in turn, phosphorylates POS-1 and inhibits the formation of stable, slow-diffusing POS-1/RNA complexes [115]. Furthermore, MEX-5 has been shown to be essential for the spatial patterning of the P granules, ribonucleoprotein organelles found in the germline cytoplasm [116, 117, 118]. P granules-proteins form germ granules by phase-transition, and this process is mediated by RNA-binding. MEX-5 competes with the P granules-protein MEG-3 and PGL-3 for access to RNA and therefore suppresses the assembly of MEG-3 and PGL-3 into phase-separated liquid droplets in the anterior. To date, a comprehensive understanding of MEX-5

functions *in vivo* has not been yet elucidated, however.

The *mex-5* gene encodes for a 468-amino acids protein with predicted low-complexity N- and C-terminal regions (see figure 3.1). The N-terminal region is rich in polyglutamine stretches that could mediate MEX-5 self-association [111]. The central domain of MEX-5, residues 268 to 341, contains two regions that are similar in sequence to the CCCH-type finger motifs that were first described in the vertebrate protein Tristetraprolin (TTP), an RNA-binding protein that regulates mRNA stability [214, 215, 216]. In TTP, each zinc finger coordinates a zinc ion by means of three cysteines and one histidine. Although several studies have characterized the RNA-binding specificity of MEX-5, insights into the structure of the TZF domain of MEX-5 are necessary to shine light onto the RNA-binding mechanisms and their contribution to the protein activity during the early stages of embryogenesis.

In this study, we present a structural and functional characterization of the TZF domain of MEX-5. We demonstrate that the N-terminal zinc finger of MEX-5 is unstructured but folds upon binding with RNA. We show that this disorder-to-order transition contributes negligibly to RNA recognition and affinity, however it is critical for the physiological activity of the protein in cytokinesis in early embryogenesis.

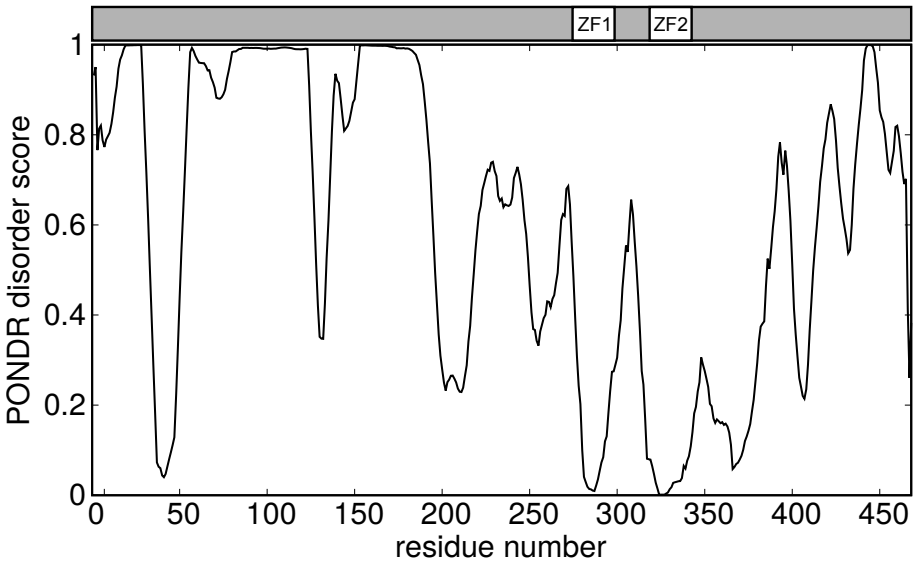


Figure 3.1: Propensities for the residues in MEX-5 to be disordered according to the PONDR algorithm.

3.3 Results

3.3.1 In the RNA-free state only the C-terminal zinc finger of MEX-5 is folded

As a first step in the characterization of the structure and dynamics of the TZF domain of MEX-5, we collected and assigned the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the MEX-$5_{268-346}$ fragment in the RNA-free form. Dispersion of the chemical shifts and analysis of the cross-peaks linewidth in a protein spectrum, in fact, provide preliminary structural information, such as presence of folded or unfolded regions [60]. We had been able to assign most of the ^{1}H-^{15}N cross-peaks to residues in the linker and in the C-terminal zinc finger (ZF2) (see figure 3.2). However, we observed that cross-peaks corresponding to residues in the N-terminal zinc finger (ZF1) were very low in intensity or broadened beyond detection. A similar result had been observed for the CCCH-type tandem zinc finger domain of the human protein Tristetraprolin (TTP) [94, 96]. For TTP, the ^{1}H-^{15}N HSQC spectrum only shows cross-peaks corresponding to its N-terminal zinc finger, whereas the peaks corresponding to the C-terminal zinc finger are broadened beyond detection. In the case of TTP, it had been showed that this behavior is due to the C-terminal zinc finger being unstructured when the protein is not bound to RNA [94, 96]. In the RNA-free form, in fact, the C-terminal zinc finger of TTP is not able to stably coordinate a zinc ion and therefore is in a molten globular state, sampling several partially disordered conformations in intermediate exchange regime [61].

To investigate the origin of the linewidth broadening of the cross-peaks corresponding to ZF1 in MEX-5, we measured the zinc-binding stoichiometry of the TZF domain of MEX-5. We collected a series of ^{1}H-^{15}N HSQC spectra at increasing concentration of zinc ions and we observed that the end-point of the titration curve occurred at one equivalent of zinc (see figure 3.3). In light of these results, we hypothesized that in a similar manner to TTP, the TZF domain of MEX-5 is also partially unstructured in

the RNA-free form.

In addition, we expressed and purify the MEX-$5_{312-346}$ construct, containing only the C-terminal zinc finger. The comparison of the ^{1}H-^{15}N HSQC spectra of the MEX-$5_{268-346}$ (TZF domain) and the MEX-$5_{312-346}$ (ZF2) constructs reveals that all the cross-peaks corresponding to ZF2 are present in both spectra, with marginal differences in chemical shifts (see figure 3.4). This result supports the idea that in both constructs the C-terminal zinc finger assumes the same three-dimensional structure and, thus, that the peptide fragment corresponding to residues 312-346 (ZF2) is able to fold and to coordinate a zinc ion when expressed alone. Moreover, we confirmed that the observed cross-peaks in the ^{1}H-^{15}N HSQC spectrum of MEX-$5_{268-346}$ correspond exclusively to residues in ZF2.

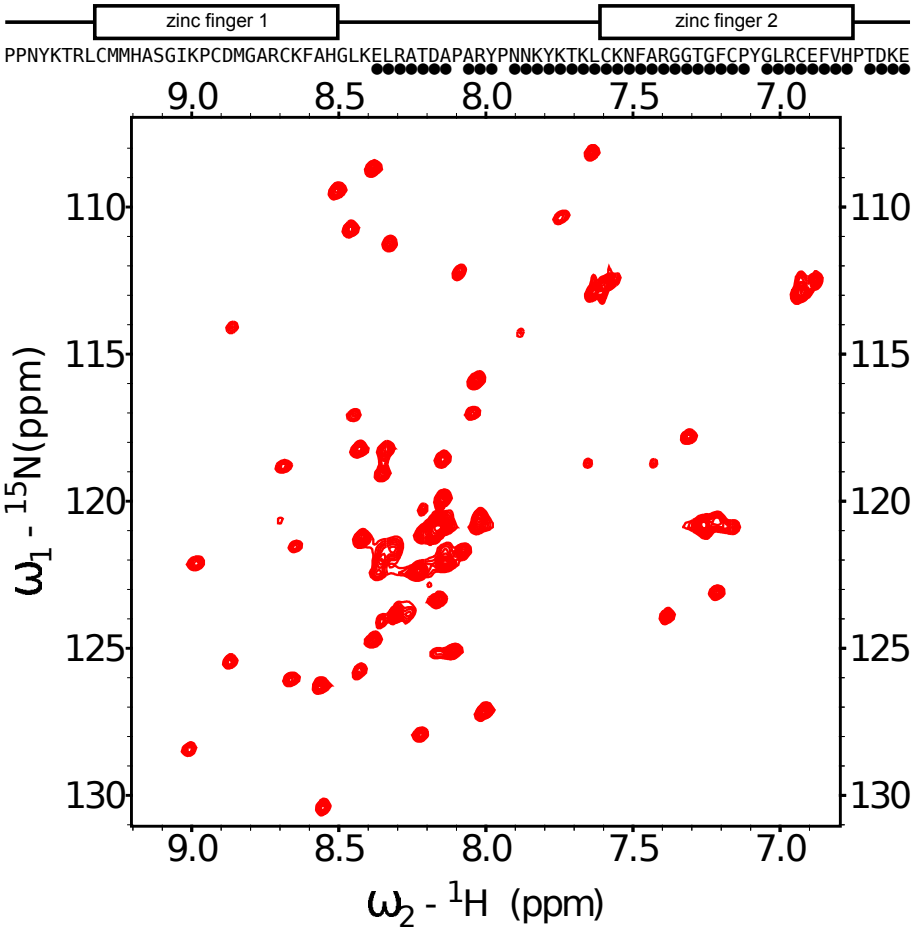


Figure 3.2: The TZF domain of MEX-5 in partially unstructured in the RNA-free state. On top, a schematic representation of the TZF domain depicts the ZFs as rectangles and the linker region as a line. The circles indicate residues along the primary sequence with an assigned cross-peak in the ^{15}N-^{1}H HSQC spectrum. On bottom, the ^{15}N-^{1}H HSQC spectrum of MEX-5 TZF domain is missing all cross-peaks from ZF1 and part of the linker.

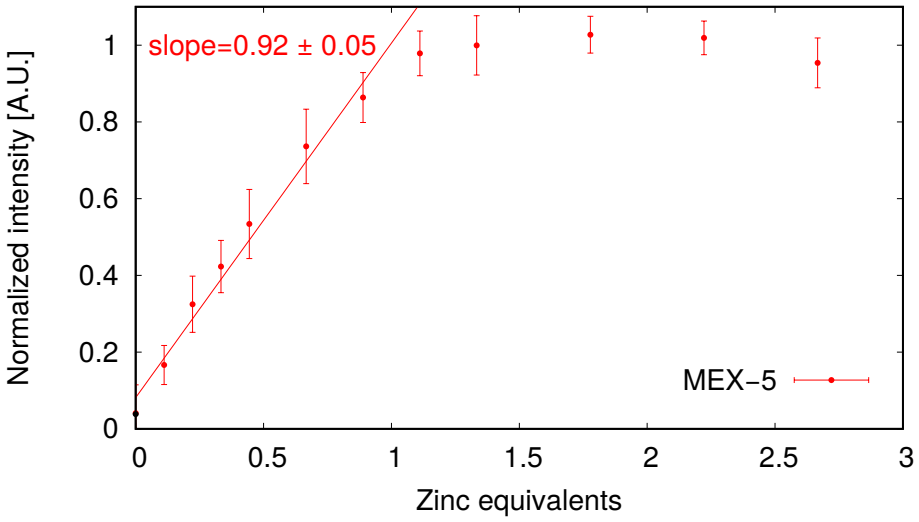


Figure 3.3: Zn^{2+} titration of the TZF domain of MEX-5, followed by NMR spectroscopy. Cross-peak intensities from the ^{15}N-^{1}H HSQC spectra are shown as a function of zinc equivalents of protein. A linear regression is applied to the first part of each curve and the resulting slope is shown in the plot.

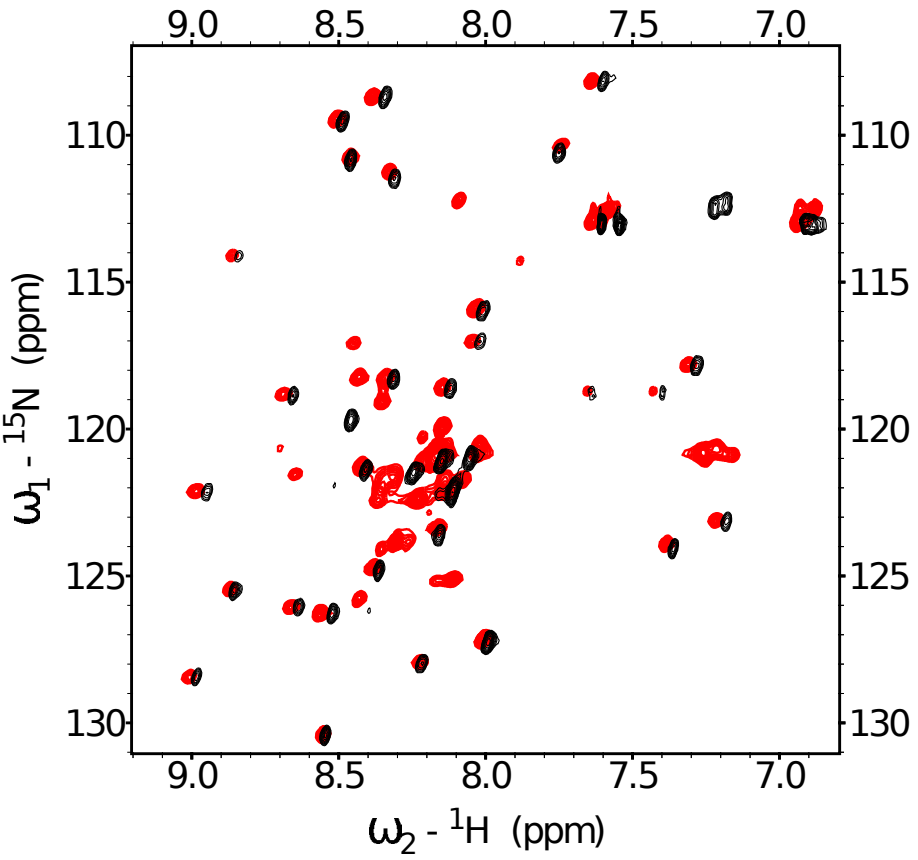


Figure 3.4: The ^{15}N-^{1}H HSQC spectrum of MEX-$5_{312-346}$ (black), containing only ZF2, has a similar number of cross-peaks with similar chemical shifts compared to the TZF domain (red).

3.3.2 Both MEX-5 zinc fingers fold upon addition of zinc and RNA

Once we established that the N-terminal zinc finger of MEX-5 is unfolded, we investigated the structure of the tandem zinc finger domain of MEX-5 bound to RNA. In 2004, Hudson et al. [90] have determined the solution structure of the TZF domains of the homolog human protein TIS11d in complex with a RNA nonamer, that to date represents the only three dimensional structure available for this class of proteins. Their results have revealed that each zinc finger of TIS11d recognizes and binds a UAUU motif. Other biochemical studies on the RNA-binding activity of CCCH-type TZF proteins [110, 96, 102, 217] have confirmed that their minimal consensus sequence for the cognate RNA molecules is spanning in length between 8 to 12 nucleotides, supporting the hypothesis that each zinc finger recognizes a motif of 3 to 4 nucleotides. For MEX-5, it has been shown that the TZF domain binds any poly-uridine stretch of at least 9 nucleotides with sub-micromolar affinity [110]. The same study also report that MEX-5 recognizes with high specificity one of the RNA target of the human homolog proteins TTP and TIS11d, the AU-rich element (ARE) in the 3′ untranslated region (3′-UTR) of the tumor necrosis factor α (TNF-α) messenger RNA (mRNA) [110].

Here, we used the RNA oligonucleotide with sequence 5′-UUUUAUUUAUUUU-3′ (ARE13) as binding partner of the TZF domain MEX-$5_{268-346}$. This sequence mimics the 3′-UTR in the TNF-α mRNA and contains the two UAUU repeats that constitute the consensus sequence of TIS11d. The apparent dissociation constant of MEX-$5_{268-346}$ for this sequence, measured using electrophoretic mobility shift assay (EMSA), is $K_d, app = 16 \pm 1$ nM. To evaluate the structural changes in the TZF of MEX-5 upon addition of RNA, we acquired a ^{1}H-^{15}N HSQC spectrum of the protein in complex with the ARE13 RNA oligonucleotide (see figure 3.5). We observed an increased number of cross-peaks in the spectrum of the bound state (66 ^{1}H-^{15}N cross-peaks for a total of 72 non-proline residues) compared to the RNA-free state (43

^{1}H-^{15}N assigned cross-peaks). The increased number of cross-peaks in the bound state spectrum suggest that residues in the N-terminal zinc finger assume a stable conformation and are not in a molten globular state. In addition, the chemical shifts of the cross-peaks in the bound state are significantly dispersed in the ^{1}H dimension and deviate from the values of residues in a random coil conformation. All together, these results show that the both the zinc fingers in the TZF domain of MEX-5 are folded, and thus coordinate a zinc ion, in presence of ARE13 RNA oligonucleotide.

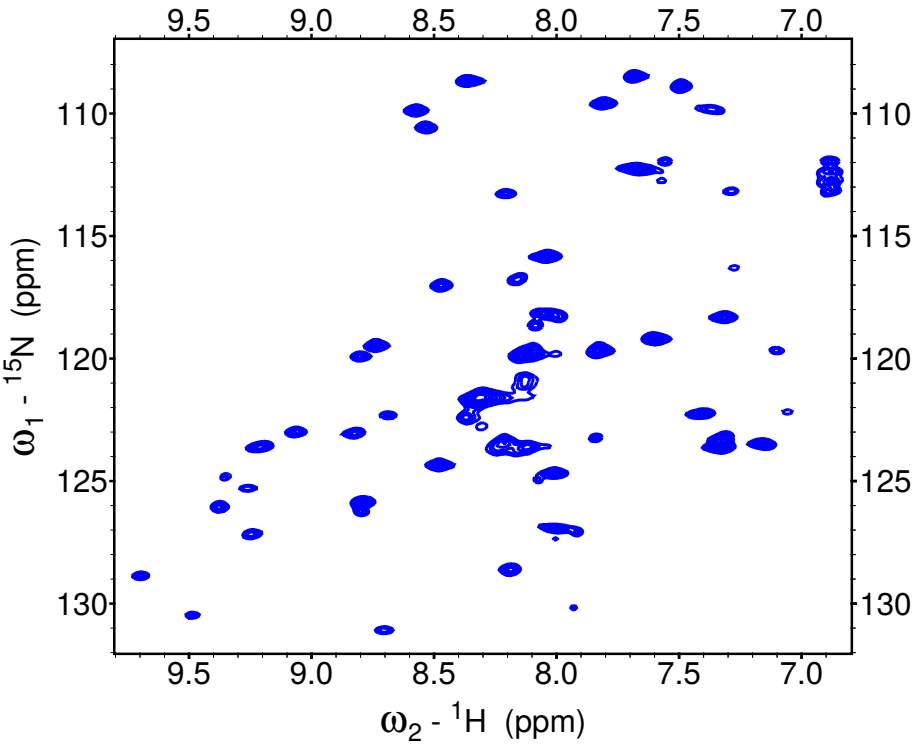


Figure 3.5: The ^{15}N-^{1}H HSQC spectrum of the TZF domain of MEX-5 bound to 5′-UUUUAUUUAUUUU-3′ RNA exhibits more cross-peaks than the RNA-free spectrum.

3.3.3 The NMR solution structure of the C-terminal zinc finger of MEX-5

To characterize the structure of the TZF domain of MEX-5, we used NMR spectroscopy. The primary sequence of the TZF domain of MEX-5 is characterized by unique features among the CCCH-type TZF protein family. The spacing between the cysteine and histidine zinc-coordinating residues, Cys_1-X(8)-Cys_2-X(5)-Cys_3-X(3)-H, is invariant in the TIS11 proteins and in many *C. elegans* homologs (see figure 3.6). In MEX-5, however, this spacing is Cys_1-X(9)-Cys_2-X(5)-Cys_3-X(3)-H in the ZF1 and Cys_1-X(10)-Cys_2-X(5)-Cys_3-X(3)-H in ZF2. The different spacing and the lack of similarity in the primary sequence with TIS11d and TTP, the CCCH-type TZF model proteins, indicate that MEX-5 zinc fingers may have a different architecture than the other members of the protein family.

Structure determination was challenging using the MEX-$5_{268-346}$ construct because of the partially unstructured nature of the domain. Being the N-terminal zinc finger partially unfolded, more than half of the residues in the TZF domain (ZF1 and the linker) are in a random coil-like extended conformation, with the corresponding cross-peaks having similar resonances. As a result, the collected spectra displayed crowded regions that made the assignment of the resonances impossible. To overcome these difficulties, for structure determination purpose we collected all the NMR experiments with the MEX-$5_{312-346}$ construct, that included only the structured C-terminal zinc finger. As discussed above, the structure of the isolated C-terminal zinc finger is not different than that of the zinc finger in the TZF domain.

The resulting ensemble of structures (root-mean-square deviation = 0.6 Å for N,Cα, C$'$, O atoms in the ordered region 320-341) (see figure 3.7 and table 3.1) showed that binding of the zinc ion, which is required to stabilize the structure of the finger, occurs through the side chains of Cys 320, Cys 331, Cys 337 and His 341. The histidine residue coordinates the zinc via its Nε2 atom. The protein structure is

further stabilized by hydrogen bonds from main chain amides to zinc-bound Cys 337 sulfur atom and by long-range backbone hydrogen bonds (see figure 3.8). In addition, the conserved aromatic side chains of Phe 323 is stacked against the side chains of His 341, suggesting a van der Waals interaction between the phenylalanine and histidine side chains that stabilizes the histidine in a rotameric state compatible with the zinc coordination geometry [218].

The polypeptide backbone adopts little regular secondary structure. Within the zinc finger there are two short 3_{10}-helix turns: Cys 320 to Ala 324, immediately after the first cysteine ligand (Cys_1), and Tyr 333 to Cys 337, immediately before the third cysteine ligand (Cys_3) (see figure 3.8). The conformation of the zinc finger (residues 320-341) is well defined and relatively rigid in the refined ensemble of structure, despite the flexibility observed for a glycine-rich loop between Cys_1 and Cys_2, specifically residues Arg 325, Gly 326, Gly 327, Thr 328 and Gly 329. Evidences of the high flexibility of the residues in the glycine-rich region are the lower ^{1}H-^{15}N heteronuclear NOEs (hetNOEs) and higher root-mean-square fluctuation (RMSF) values than the remaining of the domain (see figure 3.9). Comparing the C-terminal zinc fingers of MEX-5 and TIS11d, we observed a remarkable high similarity of the two backbone structures, with the exception of the region between Cys_1 and Cys_2 (see figure 3.10). In TIS11d this region contains a 6-residues long α-helix starting at Cys_1, whereas in MEX-5, after the short 3_{10}-helix, the two ends of the flexible glycine-rich loop are stabilized by long range interactions between Asn 322 at the N-terminal and Gly 329, Phe 330 at the C-terminal: the backbone carbonyl group of Asn 322 forms an hydrogen bond with the backbone amide group of Gly 329, and the side chain amide group of Asn 322 forms an hydrogen bond with the backbone carbonyl group of Gly 330 (see figure 3.11).

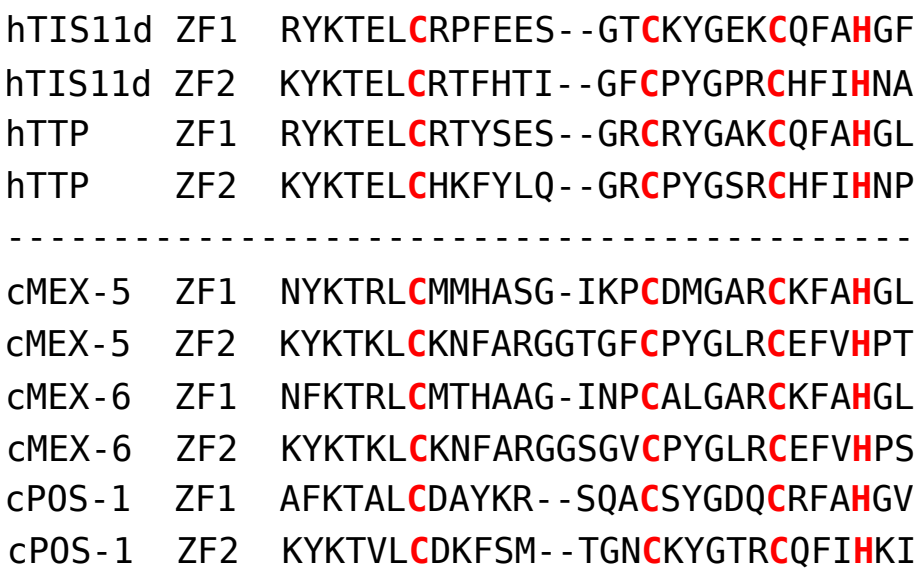


Figure 3.6: Sequence alignment of CCCH-type zinc fingers in human (h) and *C. elegans* (c). The zinc-coordinating residues are highlighted in red.

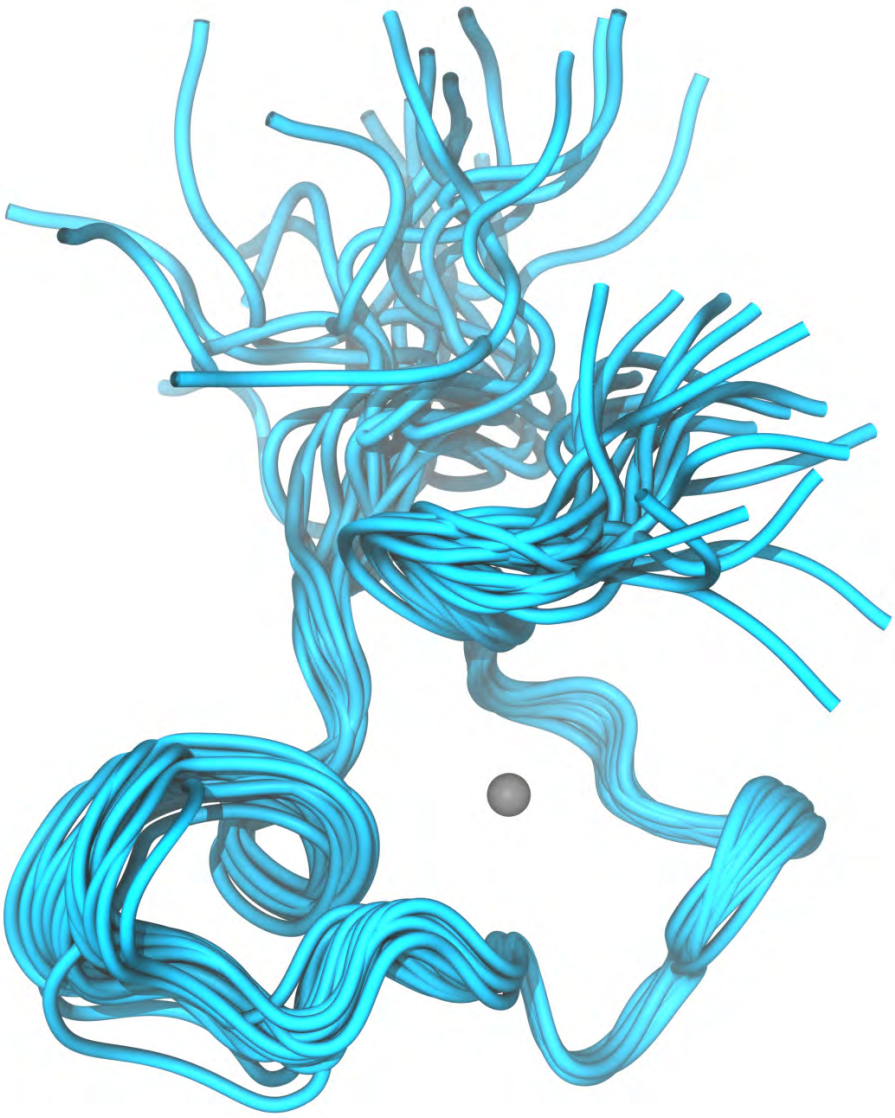


Figure 3.7: The solution structure of MEX-5 ZF2. The best 20 structures superposed on backbone heavy atoms in ordered regions of the protein are shown. Zinc is represented as a gray sphere.

Conformationally-restricting constraints	
NOE-based distance constraints:	
Total	303
intra-residue $[i = j]$	96
sequential $[\|i - j\| = 1]$	111
medium range $[1 < \|i - j\| < 5]$	40
long range $[\|i - j\| \geq 5]$	53
NOE constraints per restrained residue	9.2
Dihedral-angle constraints	43
Total number of restricting constraints	346
Total number of restricting constraints per restrained residue	10.5
Restricting long-range constraints per restrained residue	1.6
Residual constraint violations	
Average number of distance violations per structure	
0.1 - 0.2 Å	3.5
0.2 - 0.5 Å	1.1
$>$ 0.5 Å	0
average RMS of distance violation / constraint	0.05 Å
maximum distance violation	0.31 Å
Average number of dihedral angle violations per structure	
1 - 10°	3.65
$>$ 10°	0
RMS of dihedral angle violation / constraint	1.14°
Maximum dihedral angle violation	7.90°
RMSD from average coordinates	
backbone atoms	1.9 Å
heavy atoms	2.7 Å
Ramachandran statistics for ordered residues (MolProbity)	
Most favored regions	98.8%
Allowed regions	1.2%
Disallowed regions	0%
Global quality scores	Raw/Z-score
Procheck G-factor (phi / psi only)	-0.24/-0.63
Procheck G-factor (all dihedral angles)	-0.13/-0.77
Verify3D	0.20/-4.17
ProsaII	0.13/-2.15
MolProbity clash	12.43/-0.61

Table 3.1: Summary of NMR and structural statistics generated using PSVS 1.5 [219].

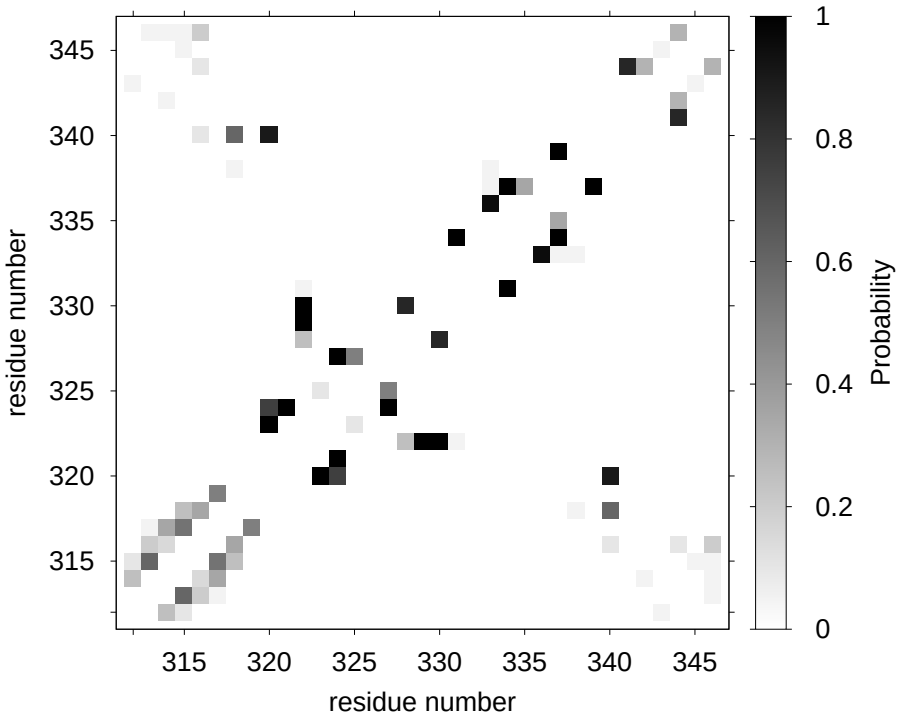


Figure 3.8: Hydrogen bonding in MEX-5 ZF2. The probability map of hydrogen bonds between two residues for the best 20 structures is shown.

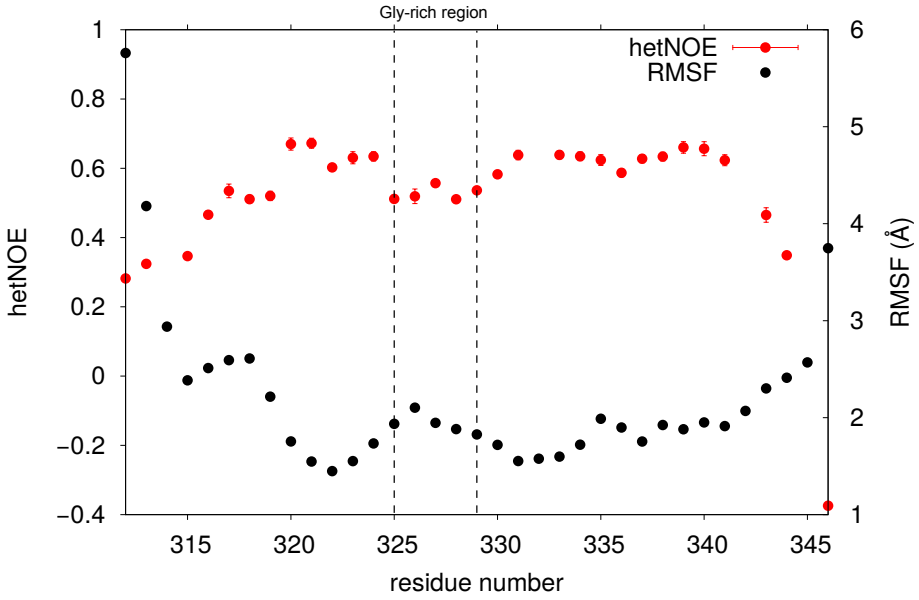


Figure 3.9: Measure of the backbone flexibility within MEX-5 ZF2. The ^{15}N-^{1}H heteronuclear NOEs measured by NMR (red) and the RMSF within the best 20 structures, for each residue, are shown.

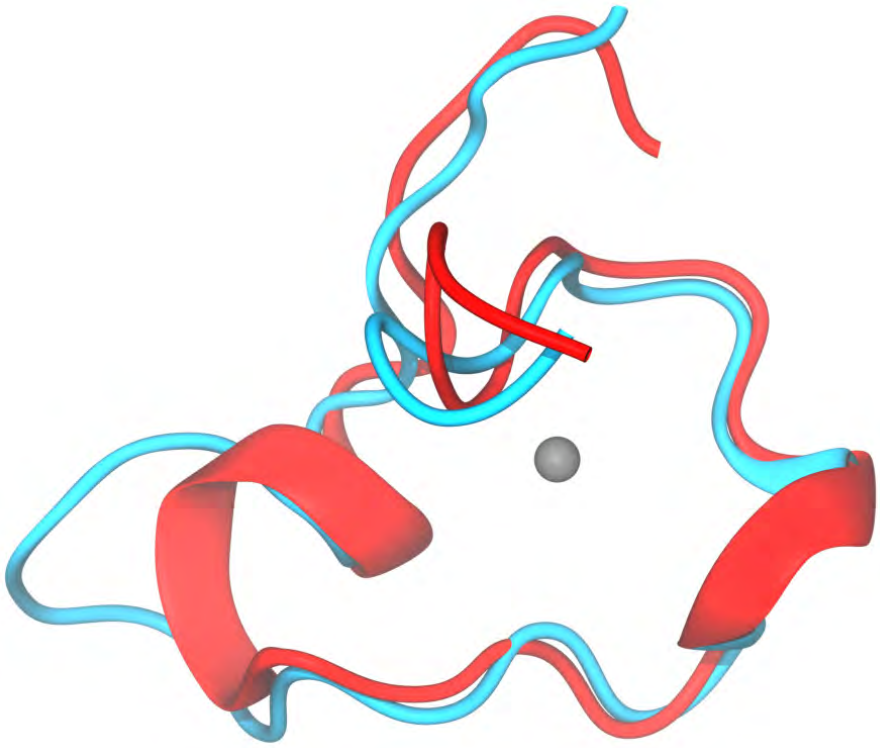


Figure 3.10: Comparison of the backbone structure between TIS11d ZF2 and MEX-5 ZF2. The lowest energy structures of the ZF2 of TIS11d (red) and MEX-5 (cyan) are shown superimposed. Zinc is represented as a gray sphere. The α-helices are represented as ribbons.

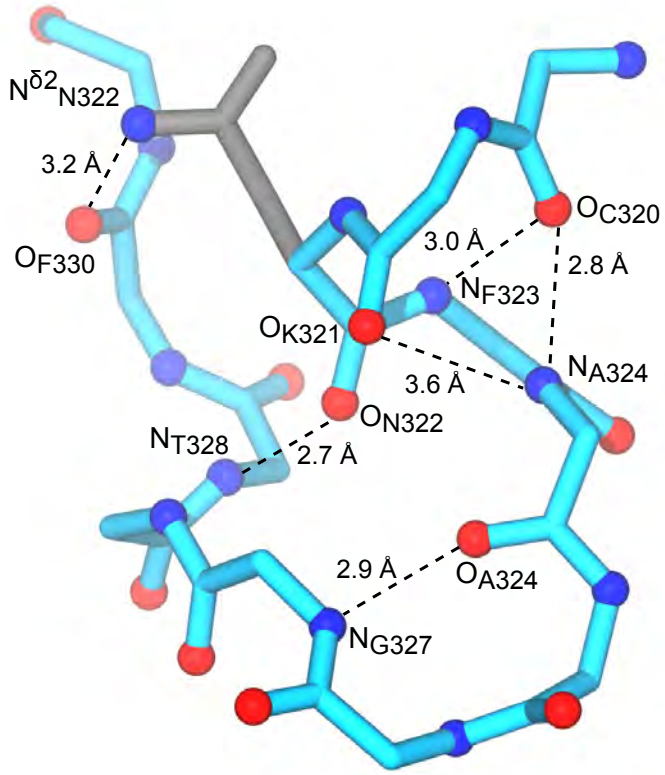


Figure 3.11: Hydrogen bonding within the Cys_1-Cys_2 region of MEX-5 ZF2. The backbone structure is depicted in cyan, oxygen atoms in red, nitrogen atoms in blue, and the side chain of Asn 322 in gray. The atoms forming H-bonds are indicated in the figure.

3.3.4 The stacking between His and Phe aromatic rings stabilizes Zn^{2+}-coordination in ZF2

To evaluate the structural differences between the two zinc fingers of MEX-5 that determine their different affinity to zinc ions, we used molecular dynamics simulations. The initial structure of MEX-5 was built using homology modeling [191, 192, 193] from the solution structure of TIS11d, solved by Hudson et al. [90]. For this reason, the initial structure of MEX-5 in the MD simulations has both zinc fingers folded and coordinating Zn^{2+} ions. As discussed before, however, we showed experimentally that only ZF2 of MEX-5 can stably bind Zn^{2+} in the RNA-free state. Consistent with this result, the three collected MD trajectories of MEX-5 exhibit the loss of tetrahedral geometry in Zn^{2+}-coordination at ZF1.

A previous study on the TIS11 proteins has established that the stacking interaction between the zinc-coordinating histidine and a conserved aromatic amino acid, three position after Cys_1, is crucial to maintain the imidazole ring of the histidine in a rotameric state compatible with zinc-binding [218]. In the TIS11 proteins the conserved aromatic moiety is in the center of a short α-helix spanning between Cys_1 and Cys_2 and, as a consequence, it is properly posed to stack against the side chain of the histidine.

In light of this result, we investigated the presence, in the collected trajectories, of such interactions in the two zinc fingers of MEX-5. We observed that in the C-terminal zinc finger His 341 stacks against Phe 323, and, thus, the side chain of the histidine is constrained in the state characterized by the dihedral angle χ_2 centered at 180° (see figure 3.12). In this conformation the zinc-coordinating residues properly and stably coordinate the zinc ion with a tetrahedral geometry. On the contrary, our simulations showed that in the N-terminal zinc finger of MEX-5 the stacking between His 279, the homologous residue of Phe 323 in ZF1, and His 296 was not maintained along the trajectories and as a result, His 296 samples multiple rotameric states, as

shown in figure 3.12). Therefore, the geometry of the zinc coordinating residues deviates from a perfect tetrahedron and in two trajectories we observed the displacement of at least one cysteine from the zinc-binding site.

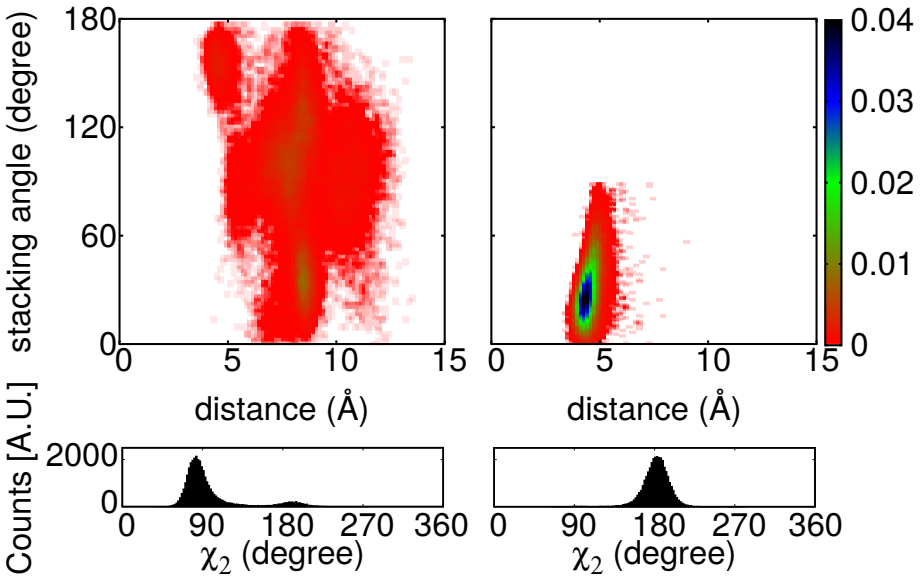


Figure 3.12: Aromatic side chains stacking in ZF2 constraints χ_2=180° of His 341. Top: Probability density distributions of the stacking angle and distance between the aromatic rings are shown for H279 and H296 in MEX-5 ZF1 (left) and for F323 and H341 in MEX-5 ZF2 (right). The stacking angle was calculated as the angle between the normals of the two aromatic rings (the planes for the side chains are defined by atoms C_{δ_2}, C_{ε_1} and N_{ε_2} for histidine and C_ζ, C_{ε_2}, C_{δ_2} for phenylalanine). The distance between the aromatic rings was calculated as the distance between the centers of mass for the heavy atoms of the two side chains. The color bars show the values of the probability density calculated for the stacking angle and distance as the number of counts normalized by the total number of observations and by the area of each bin. Configurations and distributions were extracted from three 50 ns MD trajectory.
Bottom: Probability distributions of the dihedral angle χ_2 of the zinc-coordinating histidines H296 (left) and H341 (right) are shown.

3.3.5 Rational design of a mutant of MEX-5 with a folded N-terminal zinc finger

To investigate the biological role of the disorder in the TZF domain of MEX-5, we designed mutant variants of MEX-5 where the N-terminal zinc finger is folded in absence of bound RNA. We rationalized the design of these mutants based on the solution structure of the C-terminal zinc finger described above and its difference with the structure of the human homolog TIS11d.

Our MD simulations, performed with an homology model of the TZF domain of MEX-5, revealed that in ZF1 the stacking between the side chains of residues His 279 and His 296 is not stable. The lack of this interaction destabilizes the zinc-coordination in ZF1. In the C-terminal zinc finger of MEX-5, the ligand His 341 is stabilized, by the stacking with Phe 323, in a rotameric state compatible with zinc binding [218]. In the comparison of MEX-5 to other CCCH-type zinc finger proteins, the amino acid sequence between Cys_1 and Cys_2 displays the highest degree of diversity. In particular, the spacing between the first two coordinating cysteines is 9 and 10 residues for ZF1 and ZF2 of MEX-5, respectively, and 8 residues for the other proteins of the family (see figure 3.6). Our computational studies and the primary sequence analysis indicate that the region between Cys_1 and Cys_2 alone can stabilize the structure of the zinc finger. In order to test this hypothesis, we generated a series of mutants of the TZF domain of MEX-5 where, in turn, the sequence of the Cys_1-Cys_2 fragment of ZF1 is replaced with the corresponding sequence of ZF2 (see table 3.2).

We investigated the effects of the introduced mutations on the structure of ZF1 by monitoring the ^{1}H-^{15}N HSQC spectra of the variants of MEX-5. The construct MEX-5_{CX10C} was the only one that displayed a significant difference in the ^{1}H-^{15}N HSQC spectrum compared to the wild type domain (see figure 3.13 and 3.14): we observed 78 cross-peaks in the spectrum, with those corresponding to the C-terminal zinc finger

having the same chemical shifts than the wild type. The other mutants described in table 3.2 produced ^{1}H-^{15}N HSQC spectra with a similar number of cross-peaks than the wild type domain. These results indicate that in the variant MEX-5$_{CX10C}$ both the zinc fingers are folded and coordinate zinc ions in the RNA-free form, while in the other variants ZF1 is unstructured. To confirm this hypothesis, we measured the zinc-binding stoichiometry of the TZF domain of MEX-5$_{CX10C}$. We collected a series of ^{1}H-^{15}N HSQC spectra at increasing concentration of zinc ions and we observed that, as expected, the end-point of the titration curve occurred at two equivalent of zinc (see figure 3.15).

MEX-5 ZF1	CMMHASG-IKPC
MEX-5 ZF2	CKNFARGGTGFC
	CMMHASGGIKPC
	CMMHASGAIKPC
	CMMHASGGIGPC
	CMMHASGGTGPC
	CMNHASGGIKPC
	CMNHASGGIGPC
MEX-5_{CX10C}	CKNFARGGTGFC

Table 3.2: Mutations in ZF1 of MEX-5, within Cys_1 and Cys_2, that mimic the sequence of the corresponding fragment of ZF2. Mutated residues are highlighted in red.

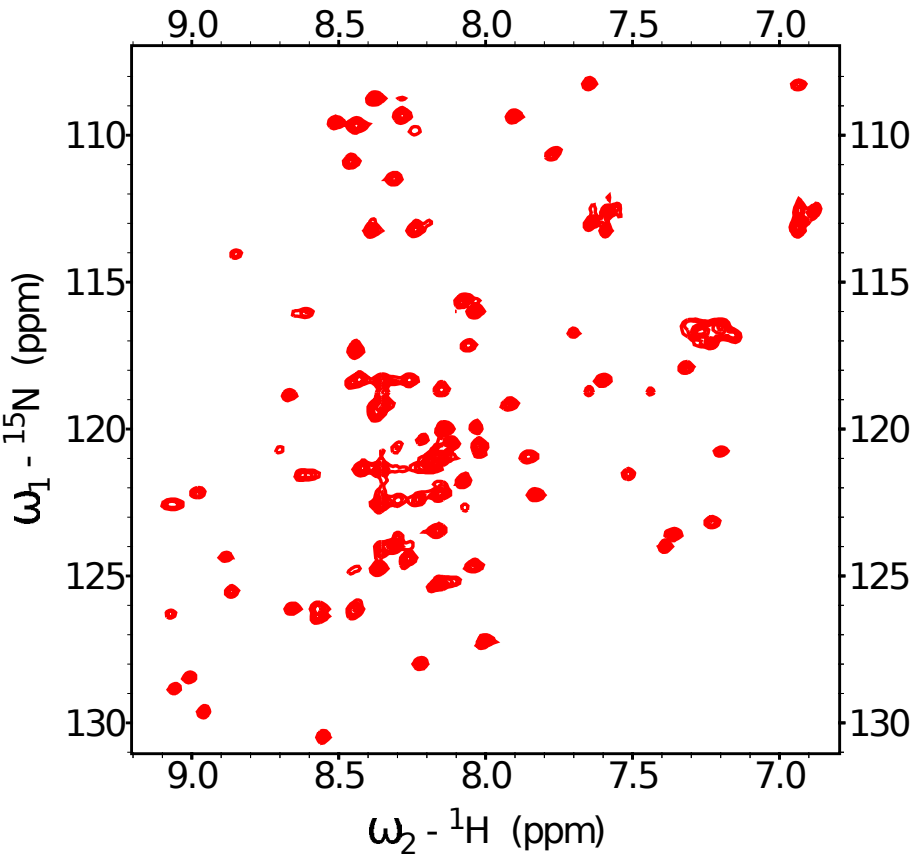


Figure 3.13: The TZF domain of MEX-$5_{CX_{10}C}$ is folded in the RNA-free state.The ^{15}N-^{1}H HSQC spectrum of MEX-$5_{CX_{10}C}$ TZF domain shows 78 cross-peaks, indicating that the domain is folded.

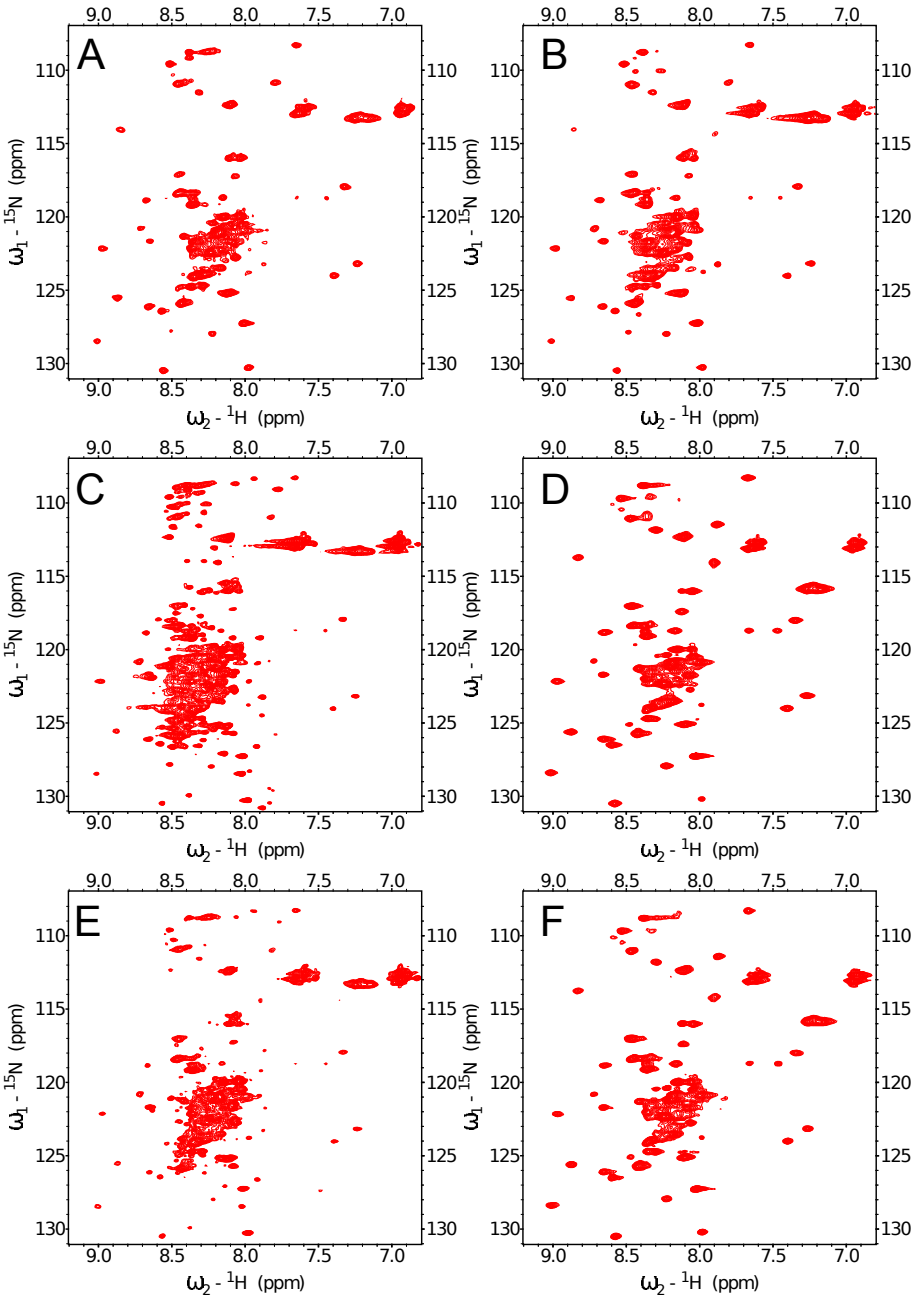


Figure 3.14: The TZF domain of MEX-5 mutants is not folded in the RNA-free state. The ^{15}N-^{1}H HSQC spectra of MEX-5 mutants from table 3.2: A: CMMHASGGIKPC, B: CMMHASGAIKPC, C: CMMHASGGIGPC, D: CMMHASGGTGPC, E: CMNHASGGIKPC, F: CMNHASGGIGPC. In spectra C and E, the cluster of overlapped peaks in the middle of the spectrum indicate aggregation.

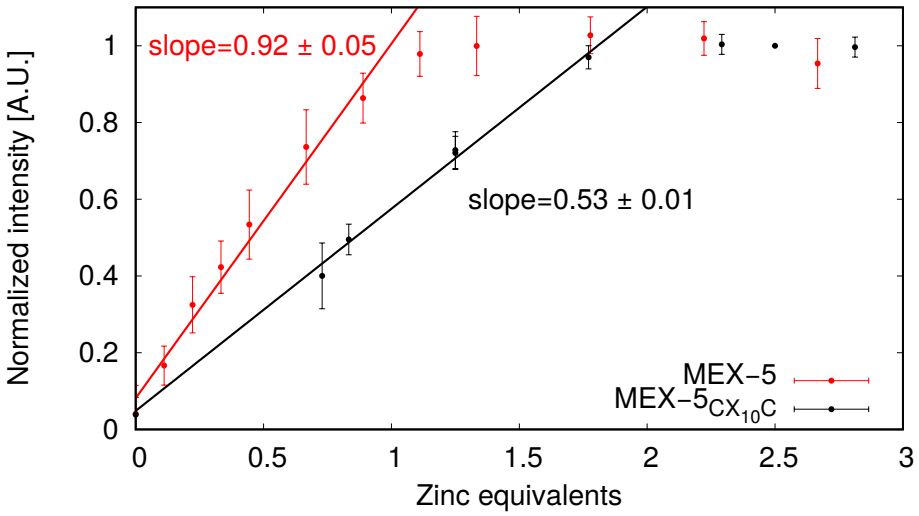
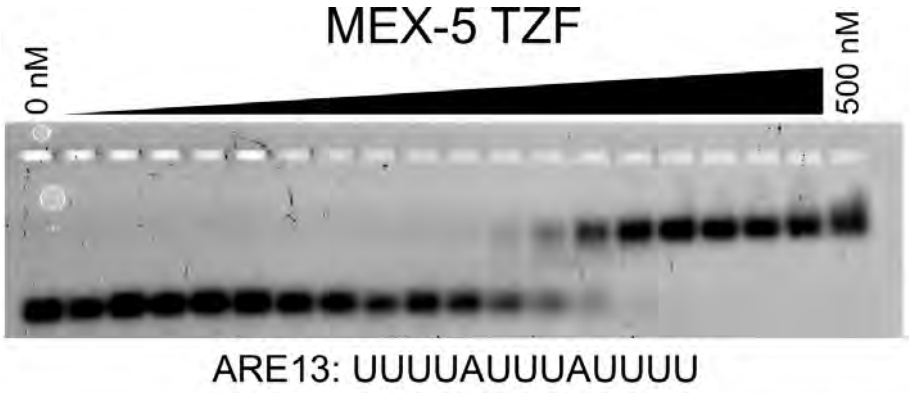


Figure 3.15: Zn^{2+} titration of the TZF domain of MEX-5 (red) and of the variant MEX-$5_{CX_{10}C}$ (black), followed by NMR spectroscopy. Cross-peak intensities from the ^{15}N-^{1}H HSQC spectra are shown as a function of zinc equivalents of protein. A linear regression is applied to the first part of each curve and the resulting slope is shown in the plot.

3.3.6 Characterization of the RNA-binding activity of the mutant MEX-5

As discussed above, the mutations introduced in the construct MEX-5_{CX10C} (see table 3.2) had the effect to stabilize the structure of ZF1, that stably coordinates zinc in the absence of RNA. To determine the effect of this structural change on the RNA-binding activity of MEX-5, we measured the apparent dissociation constant of the wild type and the mutant variant TZF domain with four different RNA oligonucleotides. The choice of the RNA target sequences was guided by the rationale that each zinc finger binds to a 4 nucleotide long U-rich motif, with higher affinity for the UAUU sequence. Thus, we measured the binding affinity of the mutated TZF domain to oligonucleotides containing in different arrangements the UAUU and UUUU elements (see figure 3.16). The results obtained for this mutant variant were compared with those of the wild type.

The results, presented in figure 3.16, show that the mutated TZF domain of MEX-5, as observed for the wild type, binds preferably to the ARE13 oligonucleotide. For all the tested sequences, we observed no significant differences in binding affinity between mutated and wild type domain ($<$ 2-fold).



RNA sequence	MEX-5 TZF	MEX-5_{CX10C} TZF
UUUUAUUUAUUUU	15±1 nM	9±1 nM
UUUUUUUUUUUUU	136±7 nM	154±4 nM
UUUUUUUUAUUUU	62±4 nM	77±6 nM
UUUUAUUUUUUUU	26±3 nM	48±5 nM

Figure 3.16: The TZF domains of MEX-5 and MEX-$5_{CX_{10}C}$ bind to the same targets with similar affinity. On top: the interaction of MEX-5 with ARE13 RNA as measured by EMSA. On bottom: The $K_{d,app}$ and the fit error of the two proteins are shown for the four RNA sequences.

3.3.7 Introducing the designed mutation into the endogenous *mex-5* locus in *Caenorhabditis elegans* by CRISPR-hr

To assess whether the lack of structure of ZF1 is important for the biological function of MEX-5, we used the CRISPR/Cas9 system to introduce the MEX-$5_{CX_{10}C}$ mutations in the endogenous *mex-5* gene. We named the resulting gene *mex-5(spr1)*. We observed that several F2 individuals exhibits the same phenotype, consisting of uterine masses (see figure 3.17), sterility and burst. In order to stably maintain the mutation *mex-5(spr1)* in heterozygotes, and thus propagate the otherwise sterile individuals, we crossed the F2 phenotypic males with VC362 hermaphrodites. The resulting animals, from now on WRM31, express pharyngeal green fluorescent protein (GFP) if heterozygotes for *mex-5(spr1)* (GFP+), but not in homozygotes (GFP-).

We observed that homozygous *mex-5(spr1)* worms lay significantly fewer embryos than heterozygous (see figure 3.18). The embryos laid by homozygous individuals fail to hatch, resulting therefore in sterility (see figure 3.18). Additionally, we characterized the survival rate of adult worms. We found that $> 50\%$ homozygotes die by bursting within 6 days post hatching, while heterozygotes exhibit longer average lifespan (see figure 3.19).

To investigate the cause of these differences between homozygous and heterozygous individuals, we analyzed the embryo morphology in WRM31 worms. Heterozygotes do not deviate from N2 worms, homozygotes develop embryo-like bodies in the uterine cavity by day 4 post hatching, however (see figure 3.20). These bodies contain DNA distribution abnormalities, as revealed by DAPI staining of the nuclei (see figure 3.20). Characterization of WRM31 homozygotes by differential interference contrast microscopy revealed polynucleated embryoid bodies at day 5 post hatching and loss of definition of the bodies by day 8, resulting in bursting of the worms (see figure 3.21).

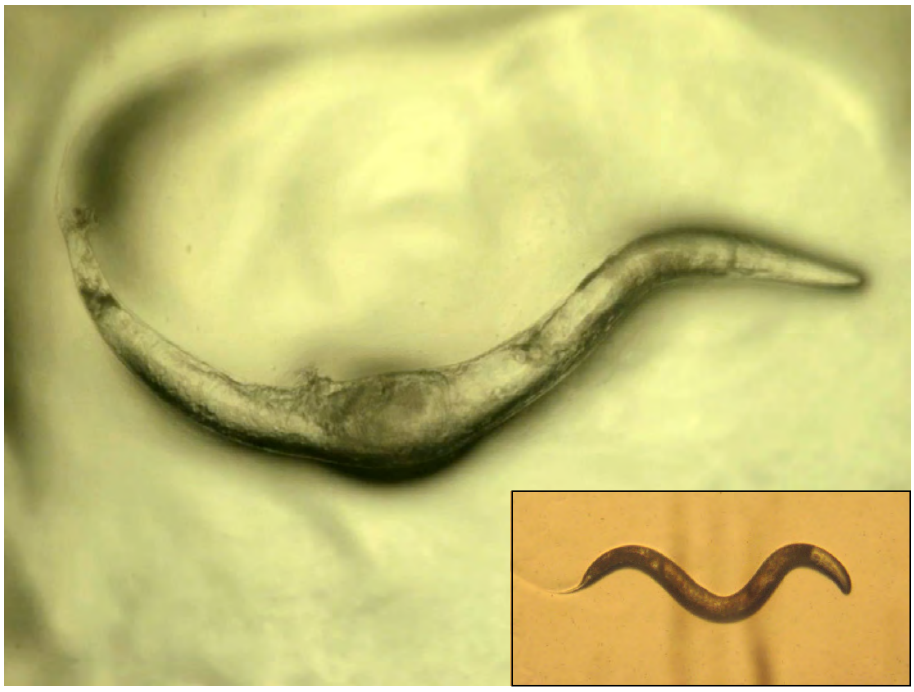


Figure 3.17: Homozygotes for *mex-5(spr1)* are sterile, develop giant uterine masses and often burst. In the picture, a homozygous WRM31 individual exhibits a giant mass in the uterin cavity. In the inset, a normal N2 individual is shown.

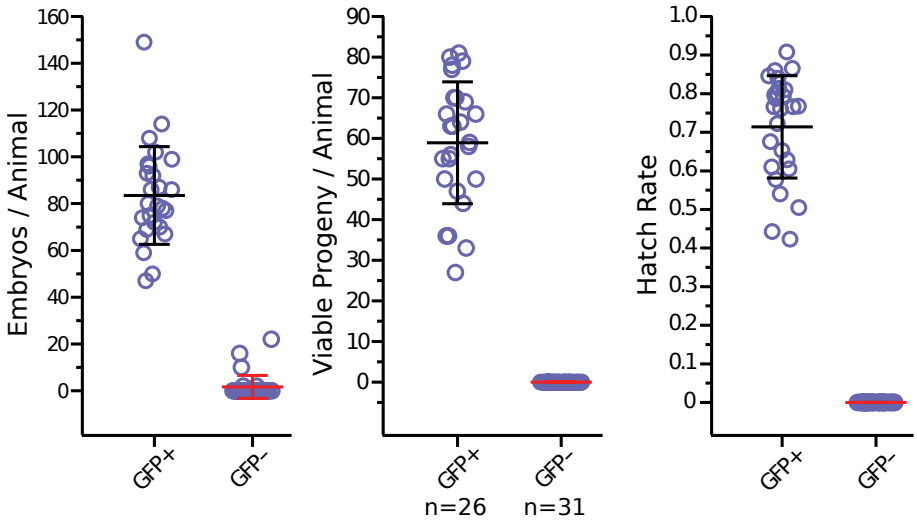


Figure 3.18: Homozygous *mex-5(spr1)* worms are sterile, but a few lay embryos that fail to hatch. Homozygous *mex-5(spr1)* worms are sterile, but a few lay embryos that fail to hatch. The the number of laid embryos per animal (left), the number of viable progeny per animal (center) and the hatch rate is shown for heterozygotes (GFP+) and homozygotes (GFP-). The number of individuals was 26 and 31 for heterozygotes and homozygotes, respectively.

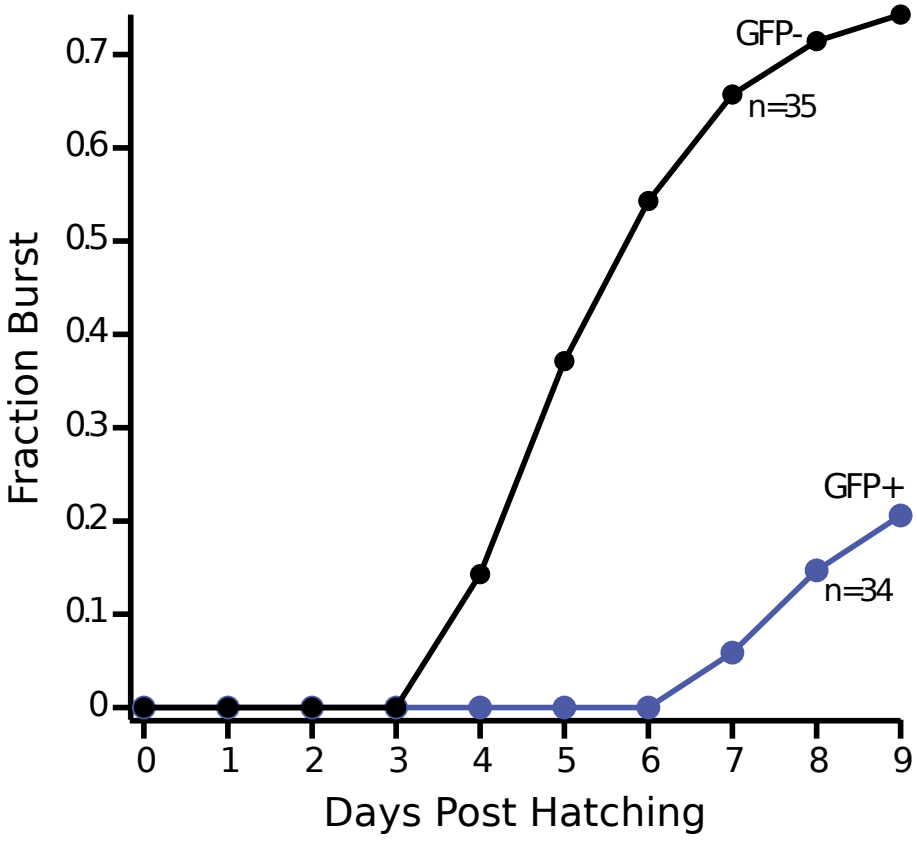


Figure 3.19: Most homozygous *mex-5(spr1)* worms die by bursting within 7 days of hatching. The fraction of homozygotes that die by bursting is displayed as a function of days post hatching.

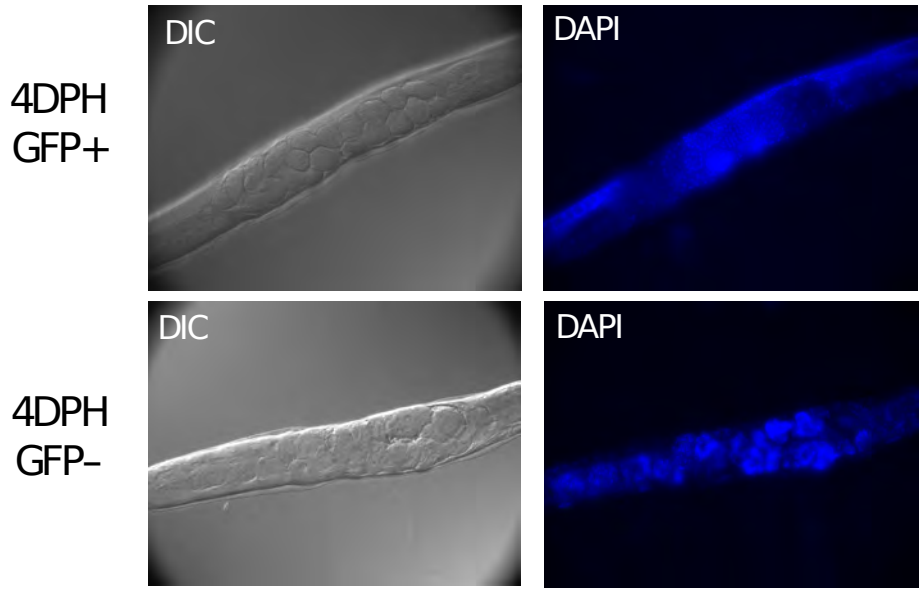


Figure 3.20: The embryo-like bodies in homozygotes *mex-5(spr1)* contain abnormal DNA distribution. Differential interference contrast microscopy images (left) and DAPI staining of the nuclei, taken at day 4 post hatching for heterozygotes (GFP+) and homozygotes (GFP-), shows mislocalization of the DNA in the uterine embryo-like bodies in homozygotes.

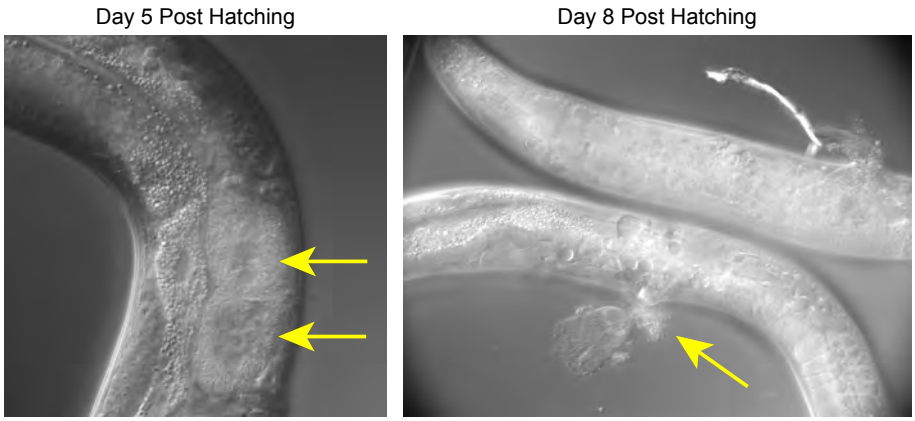


Figure 3.21: The embryo-like bodies in homozygotes *mex-5(spr1)* are polynucleated and cause bursting. Differential interference contrast microscopy images of homozygotes, taken at day 5 (left) and 8 (right) post hatching. The arrows point to polynucleated bodies in the uterus (day 5) and upon bursting (day 8).

3.4 Discussion

In this study, we characterized the structure of the TZF domain of MEX-5 (residues 268-346) to elucidate its relationship with the protein function. We demonstrated that the N-terminal zinc finger of MEX-5 is unstructured and unable to coordinate a zinc ion in the RNA-free form, but undergoes a disorder-to-order transition upon RNA-binding. Additionally, we showed that the two zinc fingers of MEX-5 fold independently: the peptide fragment containing the C-terminal zinc finger assumes the same fold of the ZF in the full-length domain context. Interestingly, by solving the solution structure of ZF2, we observed that the backbone conformation of the domain differs from the two other structures solved to this day of the members of the protein family, TIS11d [90] and POS-1 [Ertekin A., Massi F., unpublished data].

Moreover, we identified key interactions within the ZFs of MEX-5 that stabilize the zinc-binding, and based on these findings we designed a variant of the MEX-5 TZF domain that has both the ZFs structured in the RNA-free state. This variant MEX-5 TZF domain displays unchanged RNA affinity and specificity compared to the isolated wild type domain. Consequently, we introduced this variant into the endogenous *mex-5* locus in *C. elegans* using a CRISPR/Cas9-hr approach. Therefore, the resulting animals express full-length MEX-5 containing the variant TZF domain. We found that animals homozygous for the mutations are sterile and develop uterine embryoid bodies. We hypothesize that this phenotype is caused by defects in cytokinesis that lead to formation of giant polynucleated embryoid cells in the uterus.

During early stages of embryogenesis, MEX-5 contributes to the polarization of the body axes in the zygote distributing in the cytoplasm along an anterior-rich gradient. This asymmetric distribution, in turn, establishes posterior segregation of other fate determinants, POS-1 and PIE-1. MEX-5 distribution pattern in the cytoplasm has been proven to depend on its RNA-binding activity and its phosphorylation state. Additionally, MEX-5 has been shown to compete for RNA-binding with P granule-

proteins, thus disrupting their ability to form germ granules in the anterior. These protein distribution patterns in the one-cell embryo are retained upon cell divisions and result in the separation of the somatic lineage (anterior) and the germline (posterior). These observations suggest that the RNA-binding activity of MEX-5 is crucial for its function during embryogenesis.

Our findings provide new insights on the regulation of MEX-5 activity in *C. elegans*. A similar disorder-to-order transition upon RNA-binding as in MEX-5 has been observed for another protein in the CCCH-type TZF family, the human TTP, but not in the homolog TIS11d. Previous studies have determined the effect of having a fully folded RNA-binding domain on the cellular activity of TTP using a luciferase reporter assay, where luciferase was placed under the control of the TNF-α 3′ UTR [96]. Decreased reporter activity was observed when the partially unstructured RNA-binding domain of TTP was replaced with the fully structured domain of TIS11d, indicating that the increased structure is associated with higher RNA-degradation activity [96]. This result showed that folding of the RNA-binding domain is tightly coupled with TTP and TIS11d activities in the cell. In a similar manner, we hypothesize that the TZF domain of MEX-5 may have evolved to be partially structured to allow for an additional degree of regulation beside phosphorylation, that could confer functional advantages. For example, in order to bind RNA, MEX-5 requires spatially and timely availability of zinc ions in the cell to allow folding of the N-terminal zinc finger. Moreover, folding upon RNA-binding of the N-terminal zinc finger involves a conformational entropic cost that has to be compensated by a favorable enthalpic contribution to the binding energy. Folding-upon-binding is a common mechanism employed by intrinsically disordered proteins (IDPs): often IDPs are promiscuous binders that are able to interact with high specificity, but low affinity, with multiple biological targets. Although MEX-5 binds uridine-rich RNA with high affinity, it exhibits a relaxed specificity compared to other TZF proteins.

Alternatively, the mutations introduced in the variant MEX-5 TZF domain can affect other aspects of the protein function beyond the RNA-binding activity. The presence of an unstructured zinc finger in the RNA-free form, in fact, might contribute to regulate the degradation of the protein, and thus its abundance in the cytoplasm; moreover, the disordered region corresponding to the N-terminal zinc finger might act as a recognition motif for a partner protein. A caveat of our approach is that we tested the effects of the variant MEX-5 tandem zinc finger domain with a truncated form of the protein lacking the N- and the C- terminal domains. Although we haven't observed any difference in the RNA-binding specificity of the variant MEX-5 compared to the wild type form for the isolated TZF domain, we cannot exclude that the introduced mutations may affect other functions of the protein in the full-length form. In order to establish which molecular mechanism involving MEX-5 is altered by the introduced mutations, further studies of the variant TZF domain in the context of the full-length MEX-5 will be beneficial.

Previously, other studies have shown how mutations that alter the extent of structure in the free state of IDPs might affect their biological activity. Results on IκBα, an inhibitor of the transcription factor NFκB, emphasized the requirement for disorder in IκBα to efficiently dissociate NFκB from DNA [220]. Dembinski et al., in fact, showed that mutations that increase the foldedness of IκBα also decrease the IκBα-mediated NFκB-DNA dissociation rate. Their results indicate how perfectly tuned the IκBα energy landscape is for the function of dissociating NFκB-DNA from the DNA.

In conclusion, our results highlight that the relationship between the TZF domain structure and the biological function is a critical feature to elucidate known and unknown mechanisms in which MEX-5 is involved.

3.5 Material and Methods

3.5.1 Protein Expression

The RNA-binding domain of human MEX-5 (residues 268-346) and MEX-$5_{312-346}$ were amplified from pMal-MEX-5 [110] and cloned into a modified pet28 vector with a SUMO tag between BamH1 restriction site. MEX-$5_{CX_{10}C}$ was synthesized by IDT and cloned into the same vector. Mutations described in table 3.2 were generated via Quikchange mutagenesis. MEX-5 and its variants were expressed within BL21(DE3) *E. coli* competent cells. Isotopic labeling with ^{15}N and/or ^{13}C were performed by growing the cells in M9 containing 1 g of $^{15}NH_4Cl$ per liter and 2 g of ^{13}C glucose per liter, respectively. The cells were grown at 37°C to an OD_{600} of 0.8 and then induced for 4 hours with 1 mM Isopropyl β-D-1 thiogalactopyranoside (IPTG) and 0.1 mM $ZnSO_4$ at the same temperature. Harvested cells were lysed using a cell disruptor in 50 mL buffer containing 50mM Tris HCl, pH 8.0, 50 mM NaCl, and 1 EDTA free cOmplete^{TM} protease inhibitor tablet (Roche). Lysates were centrifuged at 19500 RPM for 1 hour at 4°C and passed through a 20 mL pre-packed HisTrap HP columns (GE helthcare), washed with 5 column volumes of 50 mM Tris HCl, pH 8.0, 50 mM NaCl, 20 mM imidazole, and eluted with 50 mM Tris HCl, pH 8.0, 50 mM NaCl, 350 mM imidazole. The SUMO tag was cleaved off with ULP1. The cleavage reaction was performed for 2 hours at room temperature, using a ULP1-to-protein ratio of 1:10. The protein was then passed through a 5 mL HiTRAP Q and SP column (GE Healthcare Life Sciences) pre-equilibrated with a buffer containing 50 mM Tris HCl, pH 8.0, 50 mM NaCl. Purified protein solution was buffer exchanged into 50 mM MES, pH 6.2, 100 mM KCl, 1 mM TCEP, 0.1 mM $ZnSO_4$ by dialysis and concentrated using a 3 KDa Centriprep concentrator (Millipore).

3.5.2 NMR spectroscopy

Folding of MEX-5 and its variants was monitored via NMR spectroscopy. ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) spectra were collected at 298 K on a Varian Inova spectrometer operating at 600 MHz equipped with a triple-resonance cold probe. Data processing was performed using NMRPipe [201] and Sparky [202] software.

Triple-resonance spectra were collected at 600 MHz on ^{13}C, ^{15}N-labeled protein in 92% H_2O/8% $^{2}H_2O$. Complete protein backbone 1 H, ^{13}C, and ^{15}N resonance assignments for MEX-$5_{312-346}$ were made using HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HN(CA)CO and HNCO. Aliphatic side chain assignments were made using C(CO)NH-TOCSY and H(CCO)NH-TOCSY experiments. Assignments for aromatic side chains were made from (Hβ)Cβ(CγCδ)Hδ and (Hβ)Cβ(CγCδCϵ)Hϵ spectra. A ^{15}N-edited NOESY spectrum and a ^{13}C-edited NOESY spectrum were acquired in 92% H_2O/8% $^{2}H_2O$ with a mixing time of 300 ms.

3.5.3 Structure calculations

Intramolecular protein distance restraints were derived from ^{13}C-edited and ^{15}N-edited NOESY spectra. Ambiguous distance restraints for the ZF domain and torsional restraints were generated in CYANA 2.1 [221], and the same software was used to determine initial MEX-$5_{312-346}$ structures. ϕ dihedral angle restraints were based primarily on ^{3}J HNHα coupling constants from HNHA data ($\phi = 120 \pm 45°$ if ^{3}J HNH$\alpha > 7.5$ Hz; $\phi = 50 \pm 45°$ if 3 ^{3}J HNH$\alpha < 6.0$ Hz). Additional ϕ and ψ restraints were assigned using TALOS [222], only in cases where those restraints were unambiguous and consistent with the directly determined ψ restraints. All restrained simulated annealing was done in CNS [223, 224]. The 20 unique structures with the lowest restraint violation energies were selected for analysis. During all calculations, the zinc coordination was restrained to be tetrahedral.

The protocol used here to determine the structure of the C-terminal zinc finger of MEX-5 differs from the one previously described by Hudson et al. [171] for TIS11d in several points. Inclusion of residual dipolar coupling (RDC) restraints was not required because only one zinc finger was considered in the calculation, and no additional information about the relative orientation of the two zinc fingers, as in the case of TIS11d, were necessary. Because we considered the RNA-free form of the MEX-5 C-terminal zinc finger, the zinc finger exhibited an higher degree of structural flexibility compared to the TIS11d RNA-bound domain. This feature, in particular, resulted in a lower number of structural constraints compared to TIS11d (see table 3.1). As observed by Hudson et al., in fact, the RNA provide a scaffold that stabilizes the structure of the TZF domain of TIS11d and constrains the linker between the two zinc fingers in an extended conformation [171]. As a consequence of the lack of such scaffold for the RNA-free MEX-5 C-terminal zinc finger, instead, only the region between residues 320 and 341 assumes a well defined fold, whereas the two termini are disordered (see figure 3.9).

3.5.4 MEX-5 RNA-binding domain homology model building and preparation

The structure of the ligand-free RNA-binding domain of MEX-5 (residues 268 to 346) was generated starting from the lowest energy NMR structure of TIS11d (pdb entry: 1RGO) bound to ARE (5′-UUAUUUAUU-3′) using the SWISS-MODEL Server [191, 192, 193]. The resulting structure was solvated using VMD 1.9.2 [194] in an orthorhombic water box (60x75x66 Å). Six Cl^- ions were added to the system to neutralize the charge. Three simulations of 50 ns were collected for MEX-5 as described in [218].

CHAPTER 4

CHARACTERIZATION OF THE PARTIALLY FOLDED STATES OF THE RRM2 OF TDP-43 AND THEIR SIGNIFICANCE TO ALS PATHOGENESIS

4.1 Abstract

The human protein TDP-43 is a major component of the cellular aggregates found in amyotrophic lateral sclerosis and other neurodegenerative diseases. Insoluble cytoplasmic aggregates isolated from the brain of amyotrophic lateral sclerosis and frontotemporal lobar degeneration patients contain ubiquitinated, hyperphosphorylated and N-terminally truncated TDP-43. Truncated fragments of TDP-43 identified from patient tissues contain part of the second RNA recognition motif (RRM2) and the disordered C-terminus, indicating that both domains can be involved in aggregation and toxicity. Here, we focus on RRM2. Using all-atom replica-averaged metadynamics simulations with NMR chemical shift restraints we characterized the atomic structure of non-native states of RRM2, sparsely populated under native conditions. These structures reveal the exposure to the solvent of aggregation-prone peptide regions, normally buried in the native state, supporting a role in aggregation for the partially folded states of RRM2.

4.2 Introduction

The 43 kDa human TAR DNA-binding protein (TDP-43) is a 414-amino-acid protein with two RNA recognition motifs (RRM1 and RRM2), an N-terminal dimerization domain and a C-terminal prion-like glycine-rich domain [225]. TDP-43 is a

nuclear protein that is ubiquitously expressed and whose functions include RNA splicing and transcriptional repression [226, 227]. In 2006, TDP-43 was identified as the main constituent of the cytoplasmic inclusions that are the hallmark of most forms of amyotrophic lateral sclerosis (ALS) and fronto-temporal lobar degeneration (FTLD) [122, 228]. Since then, a large number of TDP-43 mutations in familial and sporadic ALS patients have been reported (ALS Online Genetics Database) [229], thus establishing a primary role for TDP-43 in ALS pathogenesis. TDP-43 forms aggregated species *in vitro* that display high structural similarity with TDP-43 aggregates from FTLD and ALS patients [230]. Furthermore, ALS-related mutations identified in the TDP-43 gene have been reported to accelerate aggregation *in vitro* and to promote the formation of more numerous aggregates *in vivo* [230].

The cytoplasmic aggregates found in FTLD and ALS brain contain ubiquitinated, hyperphosphorylated and N-terminally cleaved TDP-43 fragments (known as C-terminal fragments) [122, 231]. TDP-43 25 kDa C-terminal fragments, identified as the pathological species in FTLD and ALS [122, 231], are generated from proteolytic cleavage within RRM2 [231, 232, 157], and consequently comprise a C-terminal part of RRM2 and the glycine-rich region. When these C-terminal fragments of TDP-43 have been expressed in cell culture, they were observed to become ubiquitinated and hyperphosphorylated and to form cytoplasmic aggregates [233, 157], recapitulating the observations made in patient-derived brain tissue [233]. Moreover, cytoplasmic inclusions formed upon ectopic expression of the C-terminal truncation product of caspase cleavage were observed to be toxic in M17 neuroblastoma cells [157]. To investigate the domain-specific contributions to neural toxicity in TDP-43, the aggregation propensity and toxicity of different TDP-43 constructs has been studied in various disease models. Studies of TDP-43 fragments with differing lengths showed that both the truncated RRM2 (residue 208-265) and the glycine-rich region contribute to formation of insoluble cytoplasmic aggregates in mouse neuroblastoma N2a cells,

with the truncated RRM2 playing a primary role [159]. In another mouse neuronal cell line (motor-neuron-derived NSC-34), the C-terminal part of RRM2 was found to be required but not sufficient for aggregation, with different TDP-43 fragments likely impairing neurite growth through different toxic mechanisms [234]. In yeast, both RRM2 and the glycine-rich region are required for aggregate formation and toxicity [158]. Taken together, these studies emphasize the importance of these two domains in the abnormal formation of cytoplasmic aggregates and in the pathogenic role of TDP-43 in FTLD and ALS.

TDP-43 RRM2 is an unusually stable domain, as shown by thermal and chemical denaturation studies [235, 159, 143]. The solution structure of TDP-43 RRM2 (PDB accession number: 1WF0, Fig. 4.4) shows a large cluster of 12 connected hydrophobic residues, isoleucines, leucines and valines (ILV) [236], in the core of the domain that likely contributes to its high stability [143]. Denaturant-induced unfolding studies have revealed the presence of an intermediate state on the RRM2 unfolding pathway. This intermediate state has a low, although not negligible, population under native conditions that increases with increasing denaturant concentration to a maximum value of approximately 80% at 4 M guanidine hydrochloride before fully unfolding [143]. This partially folded state of TDP-43, populated under native conditions, may increase the probability of cleavage leading to the production of toxic truncated forms of TDP-43 [159, 234, 158, 233, 157, 237] or directly contribute to aggregate formation. A structural characterization at atomic resolution of this partially folded intermediate state is not yet available, but it is essential to establish its role in ALS and other neurodegenerative diseases by contributing to misfolding and aggregation of TDP-43. Molecular dynamics simulations have been successfully used to explore denatured states and folding intermediates of proteins [71, 238, 239, 240]. The use of advanced sampling methods and the inclusion of experimental data as structural restraints allow sampling of low populated states in shorter simulation time and provide

an accurate representation of the system [241, 242, 243, 244, 245, 246, 247, 248, 249]. Here, we used these approaches to characterize the structure of the partially folded state of TDP-43 RRM2.

4.3 Results and Discussion

4.3.1 Free energy landscape of the RRM2 of TDP-43 in 6 M urea reveals six energy basins

We used replica-averaged metadynamics (RAM) simulations [250, 251] with backbone chemical shift restraints and bias-exchange metadynamics to characterize the partially folded state of TDP-43 RRM2. The bias-exchange metadynamics approach employed in our simulations consisted of performing four metadynamics simulations in parallel on different replicas of the system, with each simulation biasing a different collective variable (CV), as was previously described [71]. At each timestep, chemical shift restraints from experimental data collected for Cα, Cβ, N and H_N for the intermediate state at 6 M urea concentration (BMRB deposition number 27549) were applied to the ensemble of the four replicas. Mackness et al. [252] have shown that at 6 M concentration of urea the population of this partially folded state is similar to that of the native state, approximately 50%. In this study we characterize the structure of this partially folded state of TDP-43 RRM2 that is present under native conditions but sampled with low frequency. The application of the chemical shift restraints collected for the intermediate state at 6 M urea increases the frequency of sampling of this partially folded state, but the free energy landscape calculated from the simulations does not correspond to what would be observed under native conditions.

The use of chemical denaturant in general and urea in particular, is a common approach to investigate protein stability and the mechanism of unfolding [253]. An earlier study of the unfolding of chymotrypsin inhibitor 2 has shown that MD simu-

lations performed in presence of 8 M urea at 60 °C or in water at 125 °C (above the protein melting point) display the same unfolding mechanism [254]. The two simulations collected under different unfolding conditions were reported to yield similar partially folded intermediate states, sharing the same amount of secondary structure. A greater protein nonpolar solvent-accessible surface area, however, was reported for the conformations collected in the urea simulations than in water [254]. It is reasonable, therefore, to expect that the structural features of conformations obtained from our RAM simulations will be representative of the ensemble of structures observed in aqueous solution, although the relative statistical weight of each state will not be accurate.

In another study, Mielke et al. [255] showed that reference random coil chemical shifts collected in 8 M urea for $^{13}C\alpha$ and $^{13}C\beta$ have modest to negligible differences to those collected in aqueous solution. These results suggest that the chemical shift differences observed for TDP-43 at 6 M urea compared to native conditions arise mainly from conformational changes rather than solvent effects.

Appropriate choice of the collective variable is essential to guarantee proper sampling of the energy landscape [256, 257, 258]. We chose the following CVs: the total α-helical content, the total β-sheet content, the radius of gyration (R_g) and the number of hydrophobic contacts of ILV residues. The secondary structure content and R_g were chosen because fluctuations of secondary structures and volume capture the relevant structural changes sampled by a protein [71]. In addition, the amount of secondary structure and the R_g of the folding intermediate state is significantly different from that of the native state in TDP-43 RRM2 ($R_g^N = 1.41 \pm 0.01$ nm; $R_g^I = 2.34 \pm 0.04$ nm) [143, 252]. The fourth CV was chosen to be the number of hydrophobic contacts within the ILV cluster (#ILV contacts), because a large cluster of 12 ILV residues in RRM2 provides an essential core of folding stability in this domain [236, 143]. With the described approach (further details are provided in SI) we collected a total of

2500 ns of simulation time and we obtained a free energy landscape within a statistical uncertainty < 2 kJ/mol for free energies up to 10 kJ/mol (Fig. 4.6).

The collected simulations showed an effective sampling of the conformation space, with the resulting conformations spanning a wide range of values of the CVs: α-helical and β-strand content was sampled between 0% and 20%, R_g between 1 nm and 3.1 nm and #ILV contacts between 0 and 16 (Fig. 4.7). From the trajectories we calculated the free energy landscape of TDP-43 RRM2 as a function of the four CVs (Figs. 4.1 and 4.8). The resulting free energy landscape revealed six energy basins (Figs. 4.1, 4.8 and Table 4.1).

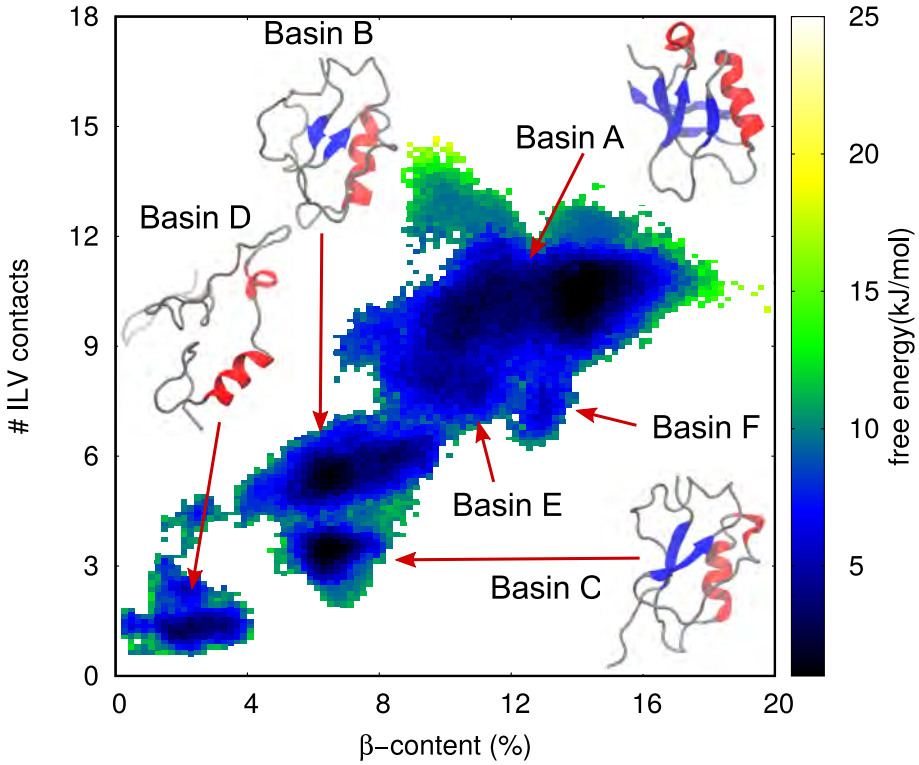


Figure 4.1: Characterization of the free energy landscape of the partially folded states of TDP-43 RRM2. Two-dimensional free energy landscape as a function of two of the four collective variables used in the replica-exchange metadynamics simulations, β-sheet content and number of hydrophobic contacts. Representative structures are shown for the basins with statistical weights greater than 10%.

4.3.2 Characterization of the residual structure and topology of the microstates of the RRM2 of TDP-43 in 6 M urea

The ensemble of conformations in the major basin, identified as A, shares most of its structural features with the native state of TDP-43 RRM2, including hydrogen bonds and contacts between ILV residues of the β-sheet and between Leu 207 in helix α_1 with hydrophobic residues in β_1, β_2 and β_4 (Fig. 4.2, 4.3 and Table 4.1). Unlike the native state, Leu 243 in helix α_2 loses interaction with β_4, but it interacts with Val 193 in β_1 (Figs. 4.2 and 4.9). This conformational change reflects the different hydrogen bonds pattern within α_2, and thus the reduced stability of the helix, compared to the native state (Fig. 4.3).

Microstates from basins B and C (Fig. 4.1 and Table 4.1) share several common structural features, including R_g and β-sheet content, and display an overall decreased structure and compactness compared to the native state and cluster A (Fig. 4.1). Both basin B and C are characterized by a structural core formed by β_1, β_3 and α_1 (Fig. 4.3), with a stable β-sheet formed by β_1 and β_3 (Figs. 4.3). In cluster C, however, β_4 and β_5 merge into an elongated strand (Fig. 4.1) that only forms marginally stable parallel pairing with β_1, not supported by any hydrophobic ILV contacts (Figs. 4.2, 4.3). A structural reorganization from β-strand to α-helix of residues 216-219, part of β_2 in the native state, is observed with higher probability in basin C than B (Figs. 4.3 and 4.10).

Basin D is the most extended and least structured cluster of microstates (Fig. 4.1). The microstates in cluster D show a high degree of structural disorder, with α_1 and a short non-native transient α-helix, formed by residues 218-221, as the only stable secondary structural elements (Figs. 4.3 and 4.10). Long-range hydrophobic contacts within the ILV cluster are limited to the interaction of Val 195 with Leu 207 and Leu 243 (Fig. 4.2). Basin E and basin F represent low populated states with no significant structural difference from cluster A (Fig. 4.11) but more relaxed packing,

as indicated by the lower #ILV contacts and larger solvent accessible surface area (SASA) (Fig. 4.12)).

In summary, the free energy landscape of TDP-43 RRM2 consists of six basins segregated in three separated regions in the CV-space (Fig. 4.1), with progressive loss of secondary structure and hydrophobic contacts, suggesting a likely path toward unfolding. The first region includes basins A, E and F, the microstates with the highest content of secondary structures and most similar to the native state. The second region, formed by basins B and C, represents conformations of RRM2 that maintain a compact structure ($R_g \sim 1.2$ nm) and several features of the topology of the native state (α_1, β_1 and β_3), but has lost several secondary structural elements (α_2, β_2, β_4 and β_5). The last region, basin D, contains the most extended and the least structured microstates of RRM2: with the exception of α_1 and a short transient α helix formed by residues 218-221 (Figs. 4.3 and 4.10), these microstates resemble a random coil conformation.

To validate the results of the RAM simulations, we calculated the backbone chemical shifts of each conformation using the software TALOS+ [222], a different algorithm than the one used in the RAM simulations. We observed good agreement with chemicals shifts measured at 6 M urea (Fig. 4.13). An R_g of 2.34 ± 0.4 nm was estimated for TDP-43 RRM2 from SAXS data collected at 6 M urea [252]. The presence of chemical denaturant in the SAXS experiments is likely to favor less compact conformations of RRM2, with a larger hydrophobic SASA, than our RAM simulations, collected in a traditional water box without denaturant [254]. The R_g values and the calculated backbone chemical shifts for basin D, that includes the most extended conformations of RRM2, are consistent with the experimental data collected at 6 M urea (Fig. 4.14). We predict these extended conformations of basin D to be significantly more populated at high concentration of denaturant.

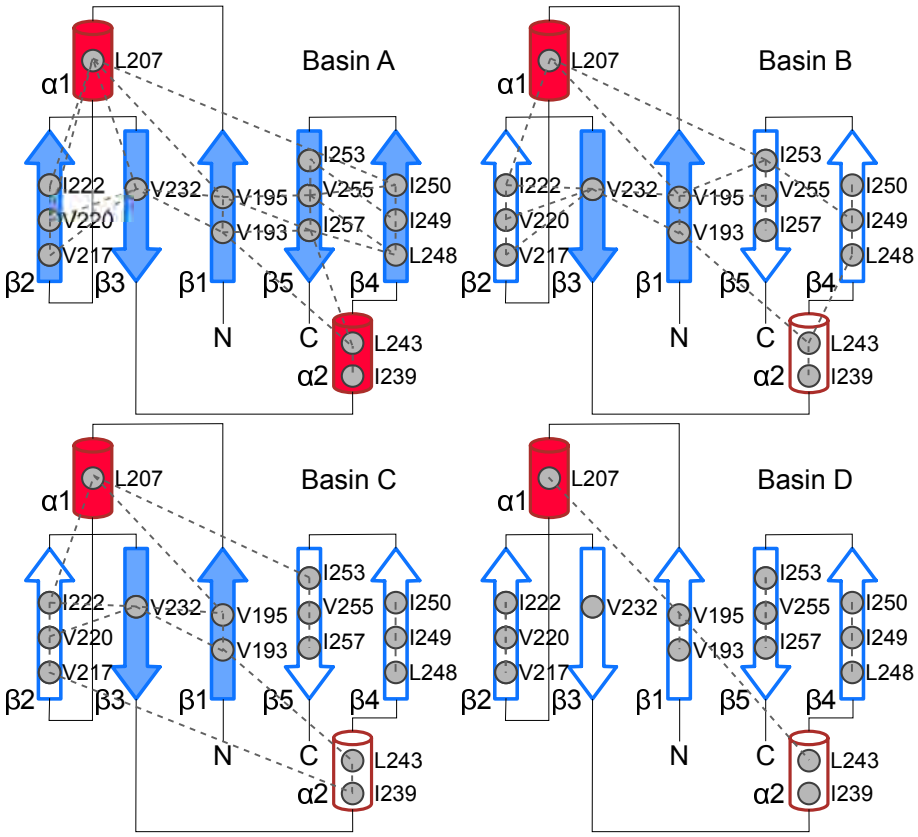


Figure 4.2: Residual structure of the microstates in the partially folded state of TDP-43 RRM2. ILV cluster contact maps for RRM2 are displayed on the secondary structure elements of the native state of the RRM2. α-helices and β-strands are depicted as red cylinders and blue arrows. Secondary structure elements are colored as solid if the structure is present in the microstate. Isoleucine, Leucine and Valine residues are shown as gray circles and ILV contacts are depicted as dashed lines.

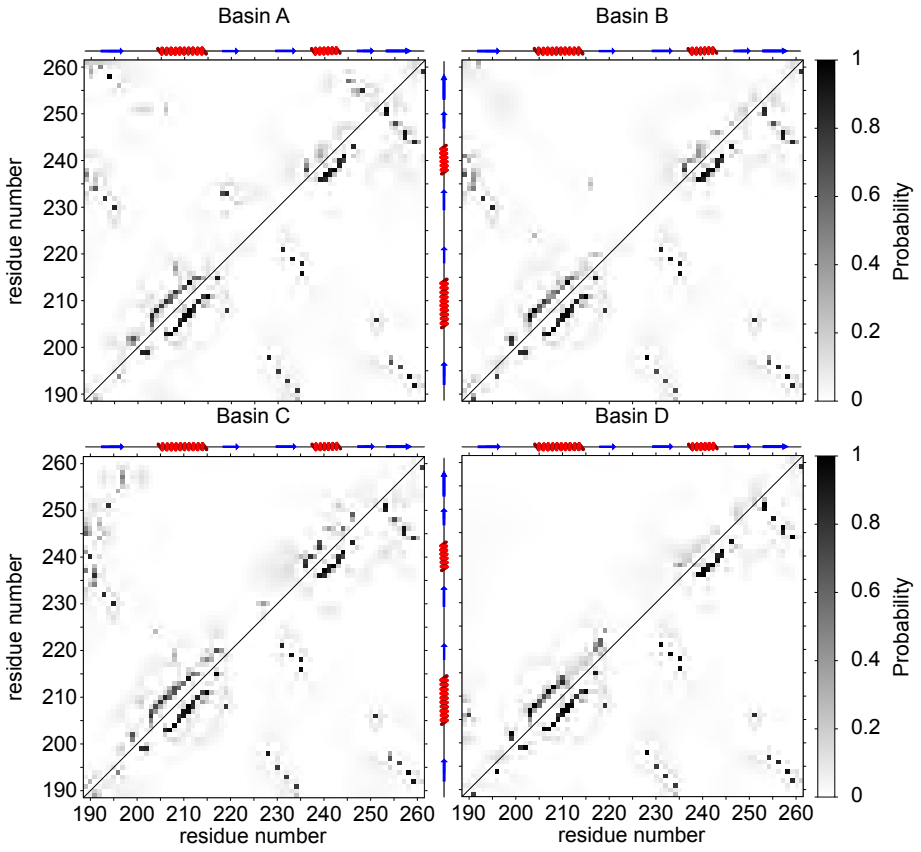


Figure 4.3: Residual structure of the microstates in the partially folded state of the TDP-43 RRM2. Comparison of the average probability of hydrogen bonds formation in the microstates corresponding to the basins shown in Figure 4.1 (above diagonal) with the native state (below diagonal). A schematic representation of the secondary structure elements of the native state is depicted on top of each map.

4.3.3 Partially folded states of the RRM2 of TDP-43 are predicted to expose to the solvent aggregation-prone β-strands β3 and β5

In order to assess a role for these partially folded states of RRM2 in TDP-43 proteinopathies, we estimated the aggregation propensity of the microstates in each basin. We employed the Aggrescan3D [259] (A3D) server to predict the aggregation propensities of each residue in the partially folded states of RRM2 corresponding to the energy basins described above. The results for the native state (Fig. 4.15 and Table 4.2) show that A3D identified only two aggregation-prone residues (A3D score > 1): Phe 221 and Ile 249. A3D predicted additional regions of RRM2 to have significant propensity to aggregate (Fig. 4.4) in the partially folded states of RRM2, Table 4.2.

In basin A two additional residues are found to be aggregation prone in β_3 and β_5 (Fig. 4.4 and Table 4.2) as a result of their increased exposure to the solvent due to the less stable pairing between the β-strands. Conformations from basins B and C are predicted to have additional aggregation-prone residues (five in basin B and nine in basic C, respectively) in regions that are part of strands β_2, β_3, β_4 and β_5 in the native state (Fig. 4.4).

The increased exposure of the hydrophobic residues in cluster B and C compared to the native state reflects the reduced β-content of these microstates, because strands β_2, β_4 and β_5 are not included in the β-sheet but are sampling a random coil conformation (Figs. 4.2 and 4.3). In accordance with the lack of secondary structures in energy basin D, conformations in this cluster are predicted to have several aggregation-prone residues, otherwise structured in the native state (Fig. 4.4). Based on the structural analysis of the microstates identified in our RAM simulations, we observed that the hydrophobic cluster centered around Val 193, Val 195 (β_1) and Val 232 (β_3) (Fig. 4.2) is critical for the stability of basin A, B, C, E and F. We predict that decreasing the hydrophobic moiety of these key residues through muta-

genesis (i.e. V193A or V195A or V232A), thus disrupting the hydrophobic cluster, will destabilize states from basins A, B, C, E and F more than the states of basin D, and will increase the aggregation propensity of the mutant variants.

The most significant feature that emerges from the A3D analysis is the increased aggregation propensity for strands β_3, β_4 and β_5 in the partially folded states of TDP-43 RRM2 relative the native state. Strands β_3 and β_5 were found to be prone to fibril formation and able to form two-dimensional sheet-like fibrils resembling that of the amyloid fibrils protofilaments in *in vitro* studies that used different synthesized peptide sequences from RRM2 [159]. Moreover, recent work has shown that peptides corresponding to RRM2 residues 247-257 (β_4 and β_5) are able to form fibrils *in vitro* through polymorphic assemblies [260]. Remarkably, the structures of the partially structured states found in this RAM simulation study shares several traits with the truncated variant of RRM2, lacking strand β_1 and part of helix α_1, found in cytoplasmic aggregates from patient tissues [231]. Indeed, the truncation of RRM2 increases the solvent accessibility of hydrophobic residues which could serve as a platform for aggregation [239].

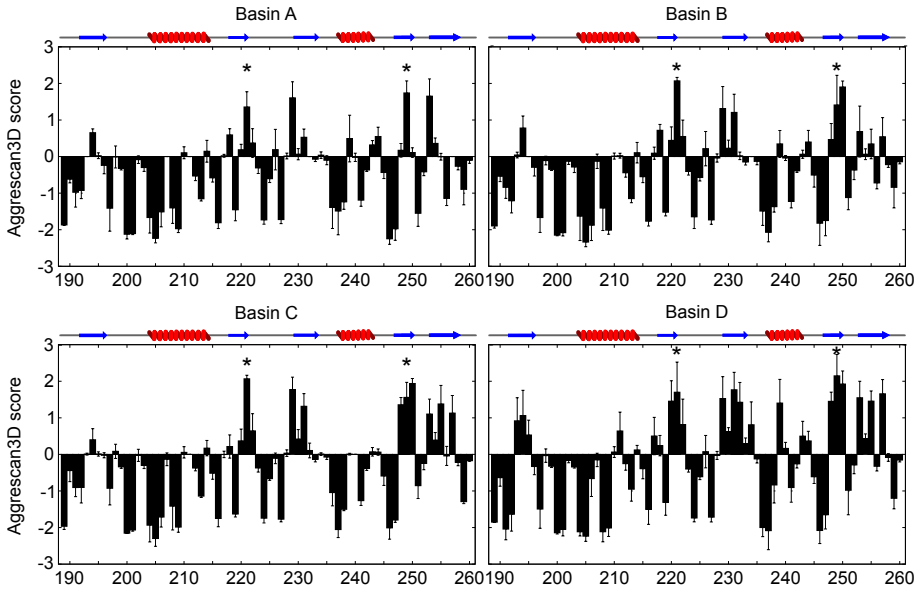


Figure 4.4: Partially folded states of the TDP-43 RRM2 exhibit increased predicted propensity to aggregation compared to the native state. The aggregation propensity of each residue in the RRM2 is shown as the Aggrescan3D score for the microstates corresponding to the basins shown in Figure 4.1. Positive values correspond to aggregation-prone residues, negative values correspond to soluble residues. Residues predicted as aggregation-prone in the native state are identified with an *. A schematic representation of the secondary structure elements of the native state are depicted on top of each plot.

4.4 Conclusions

In conclusion, we employed replica-averaged metadynamics with NMR chemical shifts as experimental restraints to characterize the structures of sparsely populated and partially folded states of TDP-43 RRM2 at equilibrium with the native state. In order to populate significantly the partially folded states, that are otherwise transient under native conditions, we used experimental data collected at high concentration of urea (6 M) and we calculated the free energy landscape of RRM2. We determined the three-dimensional structures of six ensembles of states that correspond to six distinct energy basins in the free energy landscape. Basins B, C and D (Fig. 4.1) show the loss of several elements of secondary structure (β_3, β_4, β_5 and α_2) that can expose several cleavage sites, making the protein more susceptible to truncation [261, 233]. A common feature of these partially folded conformations is the exposure to the solvent of the highly hydrophobic peptide regions located on strands β_3, β_4 and β_5, that are aggregation-prone. Previous studies had shown that short peptides carrying the sequence of strands β_3, β_4 and β_5 assembles under physiological conditions in long straight sheet-like fibrils [159, 260].

Our results provide new insights into the mechanisms underlying TDP-43 proteinopathies, proposing a role in fibrogenesis for the transient partially folded states of RRM2. These states may have a role in the function of TDP-43, for example by increasing the exposure of the nuclear export signal [262] (NES, residues 239-250)(Fig. 4.16), buried in the native state, or by facilitating interactions between the prion-like C-terminal domain and other proteins. However, upon environmental changes, such as oxidative stress [263], RNA depletion [264] or TDP-43 overexpression, the usually low-populated partially folded states of RRM2 can be sampled with higher frequency, increasing the probability of generating toxic truncated forms TDP-43 [159, 234, 158, 233, 157] and triggering self-aggregation by exposure to the solvent of the aggregation-prone strands β_3, β_4 and β_5 [159, 260]. In addition, non-native con-

formations of TDP-43 can be irreversibly sequestered in cytoplasmic fibrils, seeded by the pathogenic truncated C-terminal fragments of TDP-43 [265].

Basin	statistical weight (%)	R_g (nm)	βC (%)	αC (%)	#ILV
A	37.3	1.15±0.02	12.8±1.7	11.8±3.1	10.2±0.7
B	11.5	1.20±0.02	7.2±1.3	11.2±2.2	5.7±0.6
C	10.7	1.24±0.03	6.7±0.6	12.6±1.3	3.4±0.4
D	10.6	1.61±0.25	1.9±0.8	14.2±1.5	1.8±0.8
E	4.8	1.20±0.02	10.0±0.7	12.9±3.0	8.0±0.4
F	0.5	1.18±0.01	13.0±0.2	11.4±2.2	7.7±0.4

Table 4.1: TDP-43 RRM2 Free energy landscape

native state	
	β_2:Phe 221
	β_4:Ile 249
partially unfolded states	
basin A, E, F	β_2:Phe 221
	β_3:Phe 229
	β_4:Ile 249
	β_5:Ile 253
basin B	β_2:Phe 221
	β_3:Phe 229 and Phe 231
	β_4:Ile 249 and Ile 250
basin C	β_2:Phe 221
	β_3:Phe 229 and Phe 231
	β_4:Leu 248, Ile 249 and Ile 250
	β_5:Ile 253, Val 255 and Ile 257
basin D	β_1:Phe 194
	β_2:Val 220 and Phe 221
	β_3:Phe 229, Phe 231 and Val 232
	α_2:Ile 239
	β_4:Leu 248, Ile 249 and Ile 250
	β_5:Ile 253, Val 255 and Ile 257

Table 4.2: TDP-43 RRM2 aggregation prone residues. List of residues with an A3D score > 1 for the native state (PDB ID: 1WF0) and partially folded conformations of TDP-43 RRM2 obtained from the RAM simulations. The residues are organized according to their location in the native state.

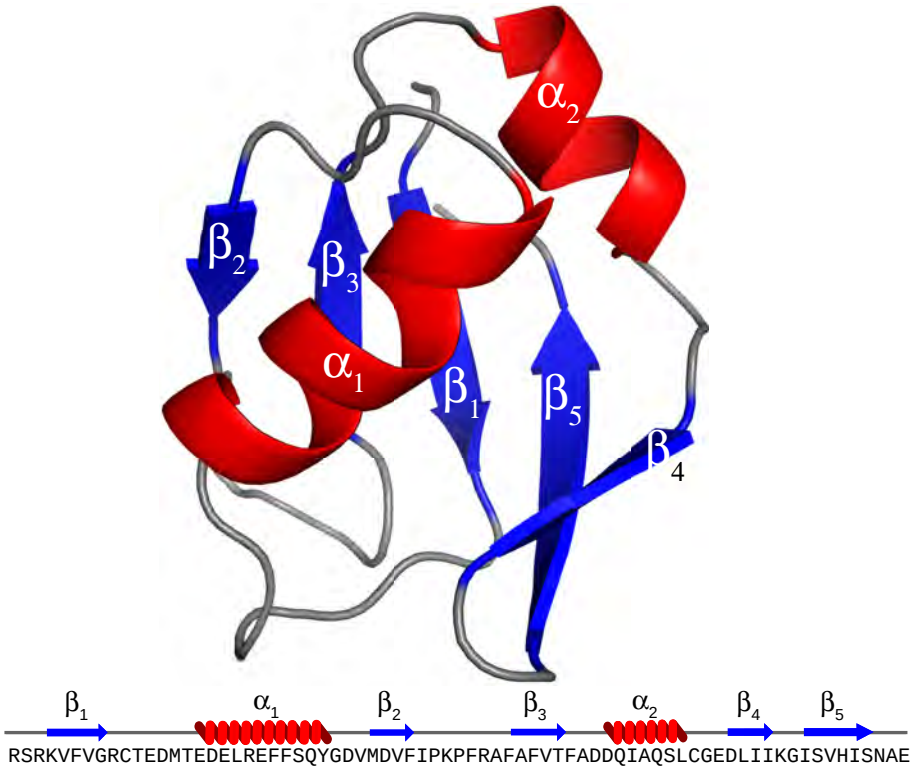


Figure 4.5: Structure of the native state of TDP-43 RRM2. Top: the structure of TDP-43 RRM2 with secondary structures depicted as red α-helices and blue β-strands. Bottom: sequence of TDP-43 RRM2 with schematic representation of the secondary structures.

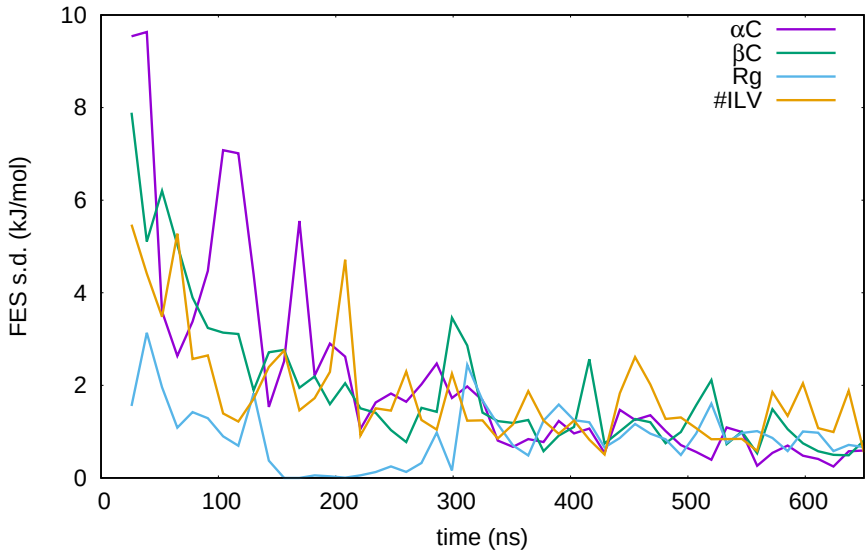
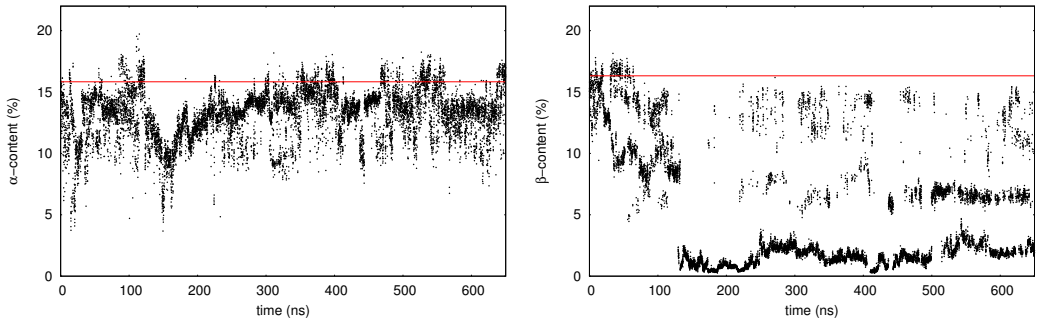
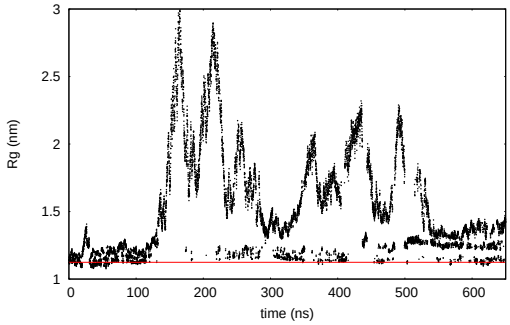


Figure 4.6: The convergence of the free energy calculations is evaluated by considering the free energy change as a function of the simulation time. The free energy change is calculated as the difference between the free energy at time t-t_0 and the free energy at time t, where t_0 is 13 ns.

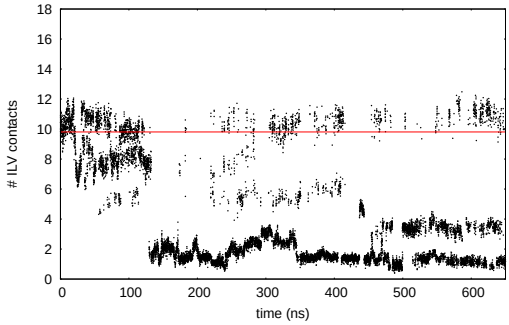


(a) CV1

(b) CV2



(c) CV3



(d) CV4

Figure 4.7: The values of the four collective variables employed in the RAM simulation as function of time are reported for one of the four replicas. A red line indicates the value of the collective variable for the native state in each plot.

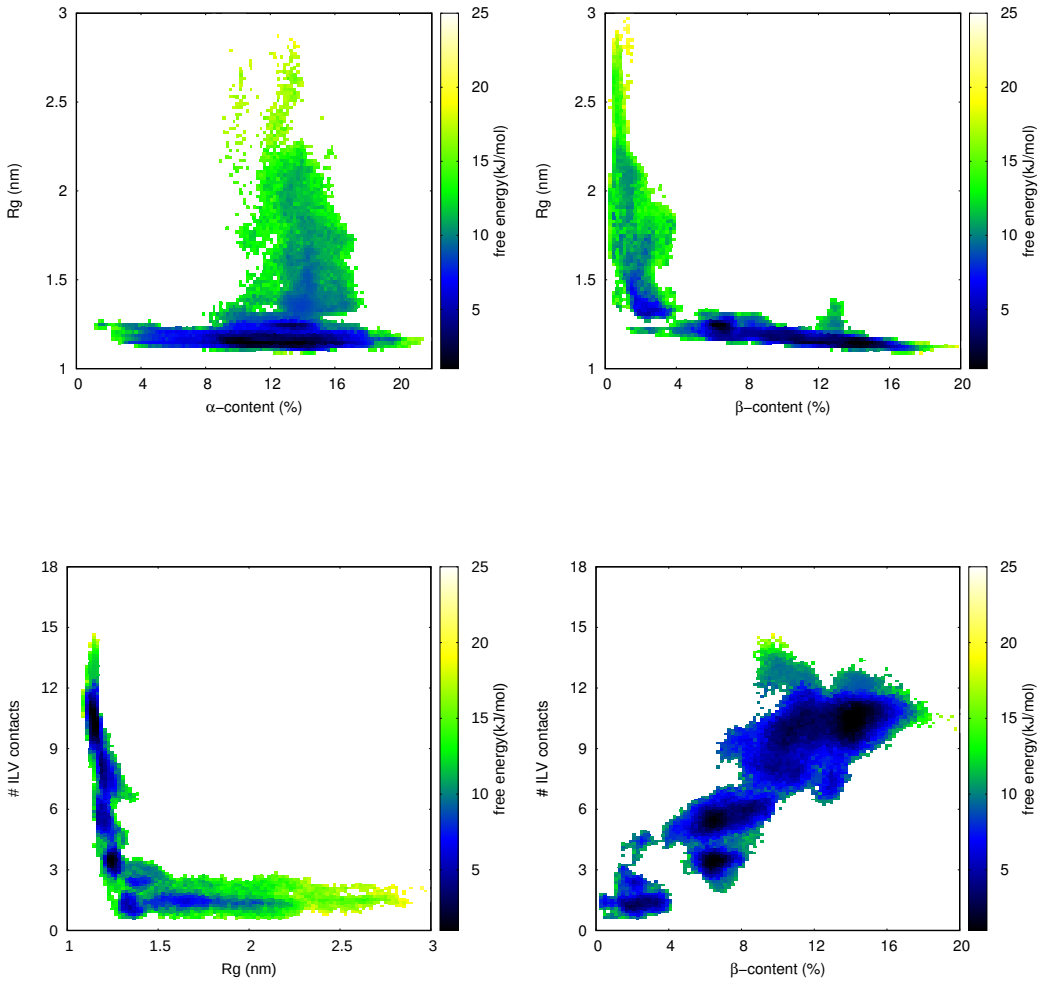


Figure 4.8: Characterization of the free energy landscape of the partially folded states of the RRM2 of TDP-43. Two-dimensional free energy landscapes as a function of two of the four collective variables used in the replica-exchange metadynamics simulations.

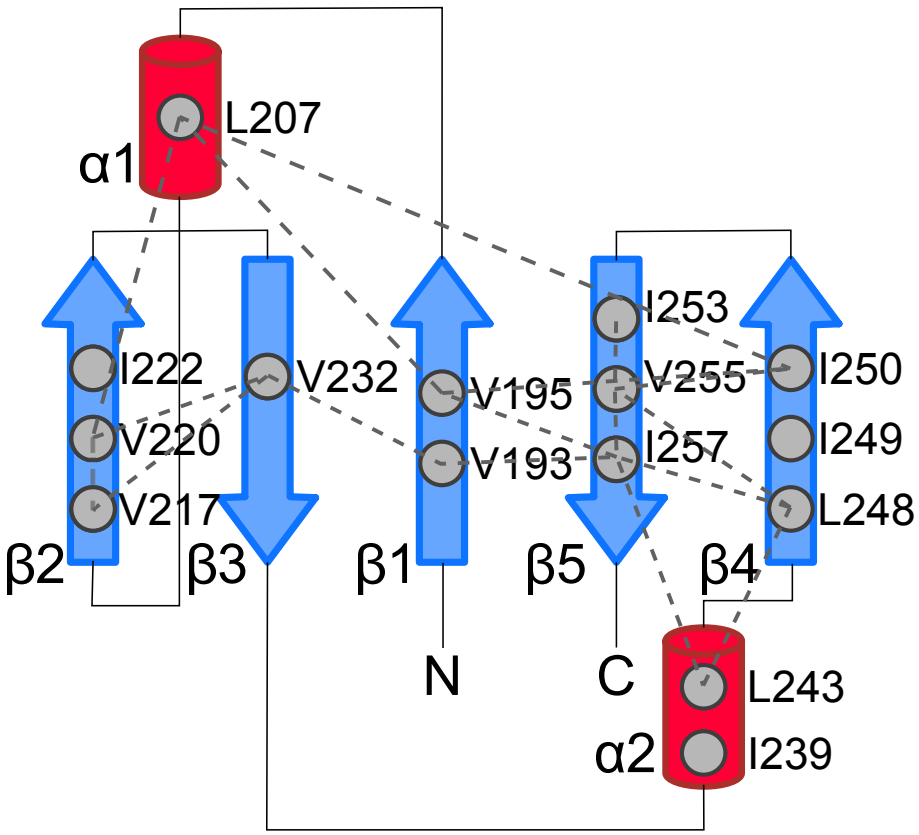


Figure 4.9: Native ILV cluster contact map for RRM2 is displayed on the secondary structure elements of the native state of the RRM2. α-helices and β-strands are depicted as red cilinders and blue arrows. Secondary structure elements are colored as solid if the structure is present in the microstate. Isoleucine, Leucine and Valine residues are shown as gray circles and ILV contacts are depicted as dashed lines.

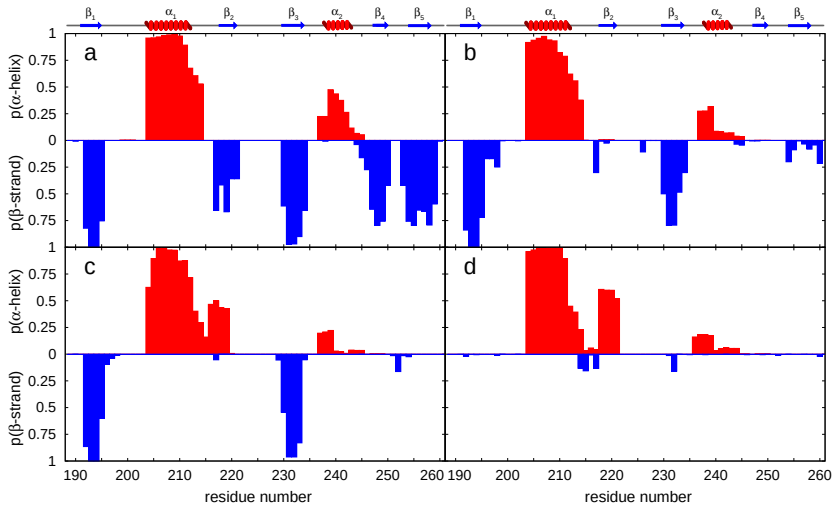
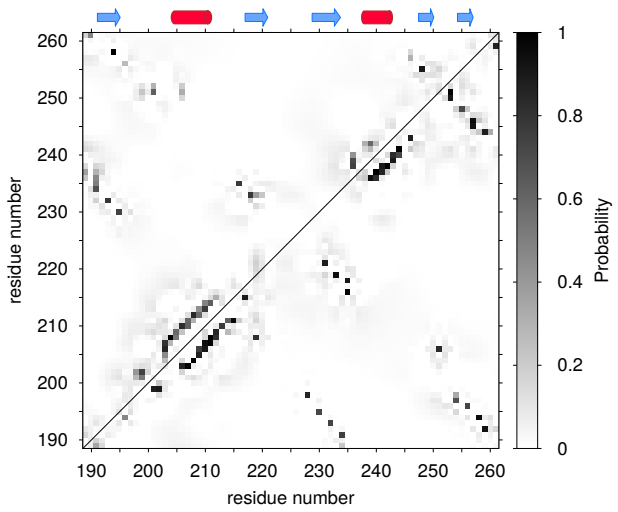
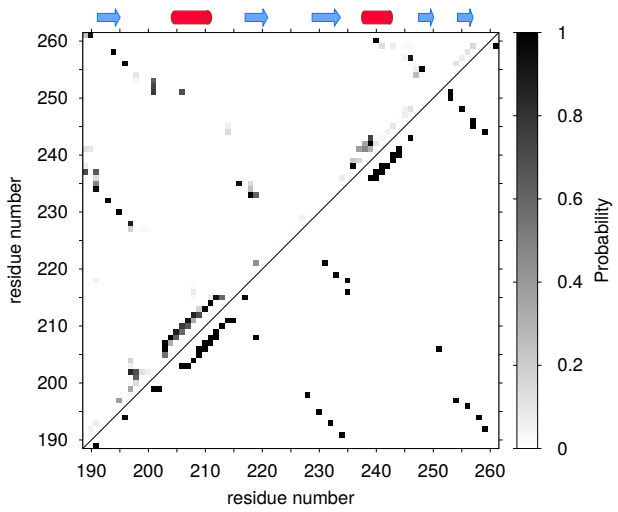


Figure 4.10: Secondary structures in the partially stable states of TDP-43 RRM2. The probability of each residue being in an α-helix or a β-strand is shown as computed by the DSSP algorithm [266] for the microstates in basin A (a), B (b), C (c) and D (d).



(a) Basin E.



(b) Basin F.

Figure 4.11: Residual structure of the microstates in the partially folded state of the RRM2 of TDP-43. Comparison of the average probability of hydrogen bonds formation in the microstates corresponding to the basins E and F (above diagonal) with the native state (below diagonal). A schematic representation of the secondary structure elements of the native state is depicted on top of each map.

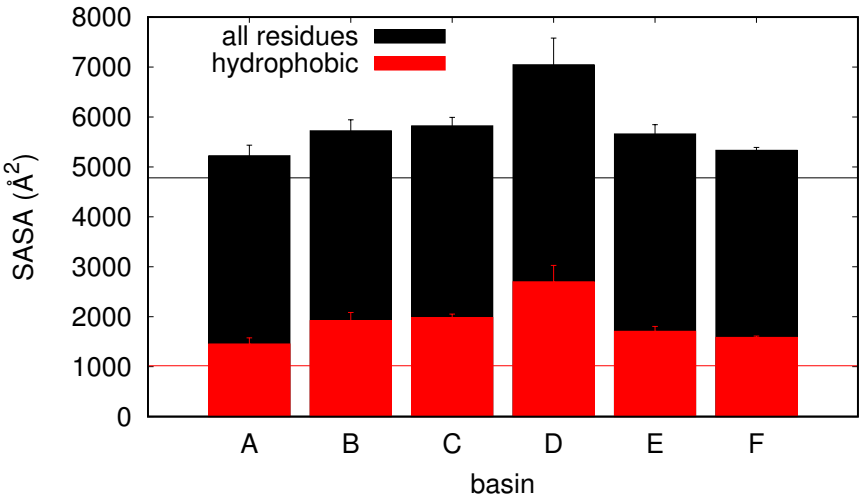


Figure 4.12: Solvent accessible surface area of the partially folded states of TDP-43 RRM2. The SASA of RRM2 (all residues) and of the hydrophobic residues (Ala, Leu, Val, Ile, Pro, Phe, Met, Trp) are shown for the microstates corresponding to the energy basins. The horizontal lines in the plot indicate the values for the native state.

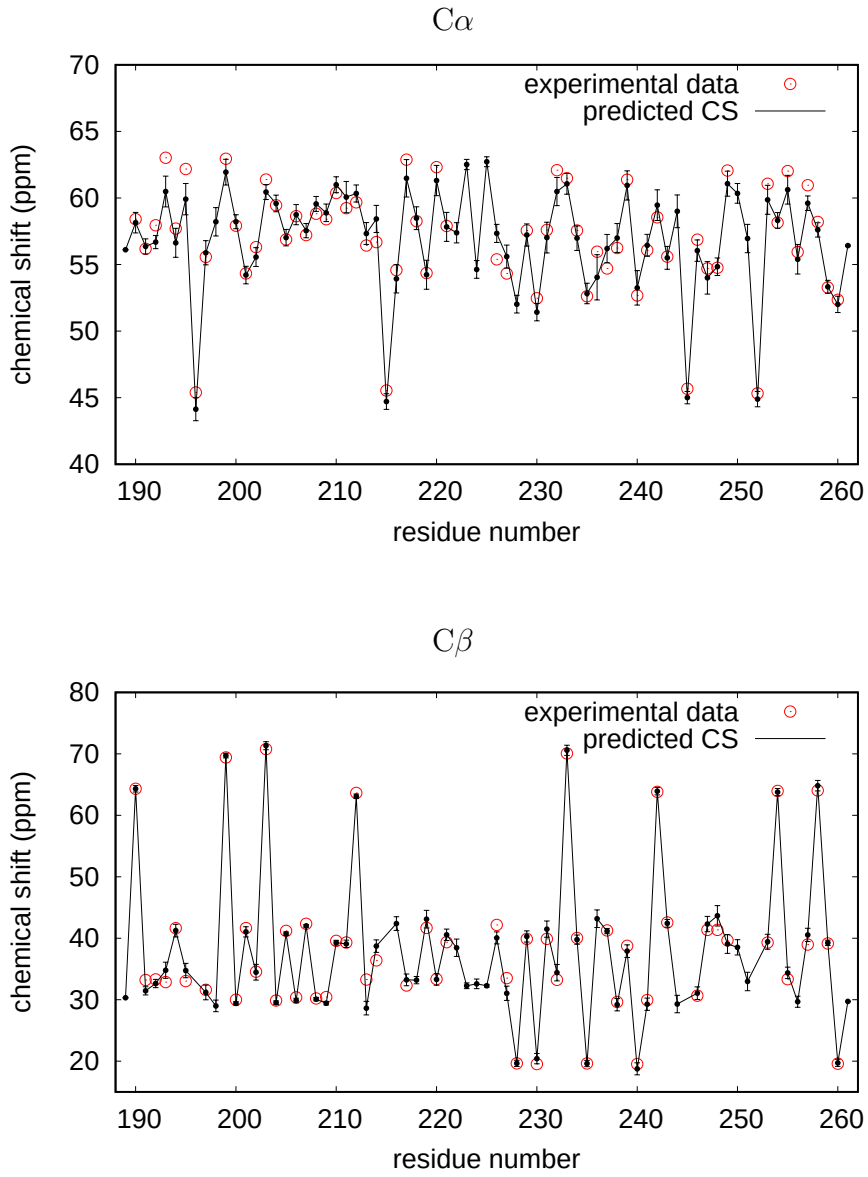


Figure 4.13: Comparison of the experimental Cα and Cβ chemical shifts with the values backcalculated using TALOS+ [222] for the microstates obtained from the RAM simulations. The values are shown as weighted average over the entire ensemble of structures in the energy landscape.

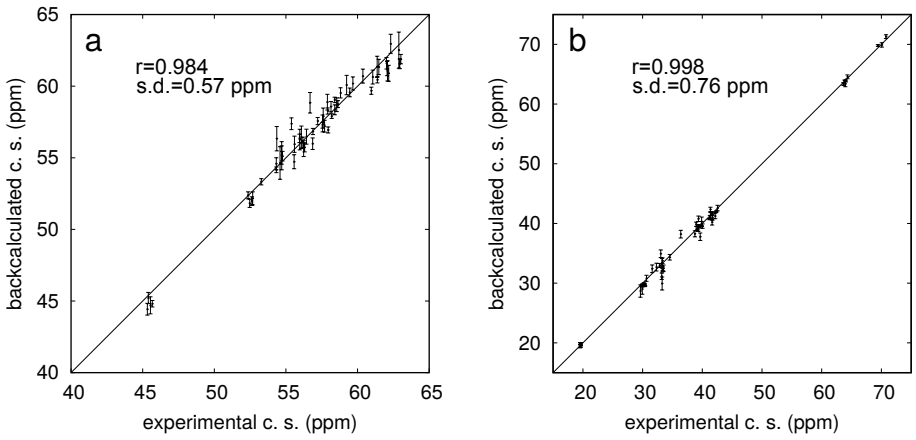


Figure 4.14: Correlation of the backcalculated Cα (a) and Cβ (b) chemical shifts for basin D with the experimental data collected at 6 M urea.

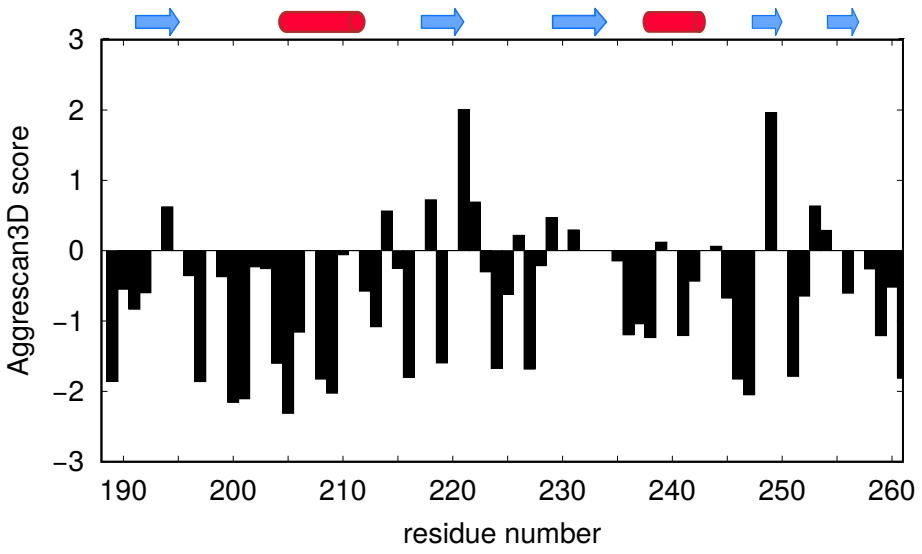


Figure 4.15: The aggregation propensity of each residue in RRM2 is shown as the Aggrescan3D [259] score for the native state. Positive values correspond to aggregation-prone residues, negative values correspond to soluble residue. A schematic representation of the secondary structure elements of the native state are depicted on top of the plot.

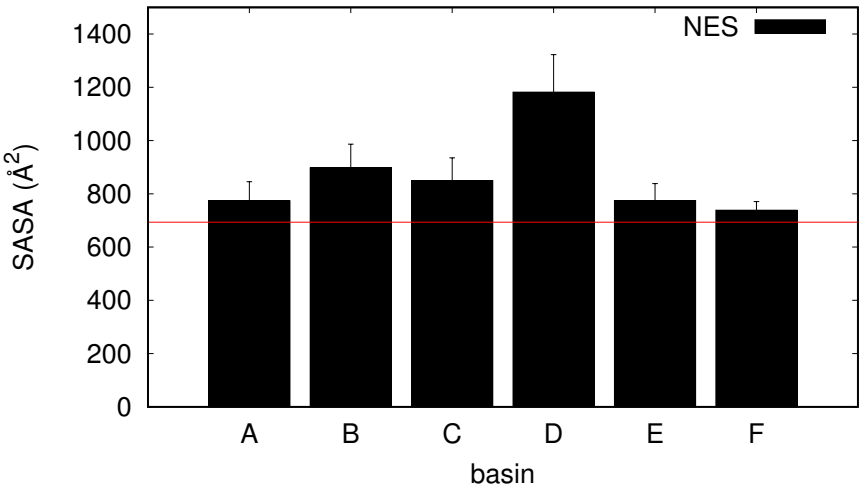


Figure 4.16: Solvent accessible surface area of the partially folded states of TDP-43 RRM2. The SASA of the nuclear export signal (NES, residues 239-250) region is shown for the microstates corresponding to the energy basins. The horizontal lines in the plot indicate the value for the native state.

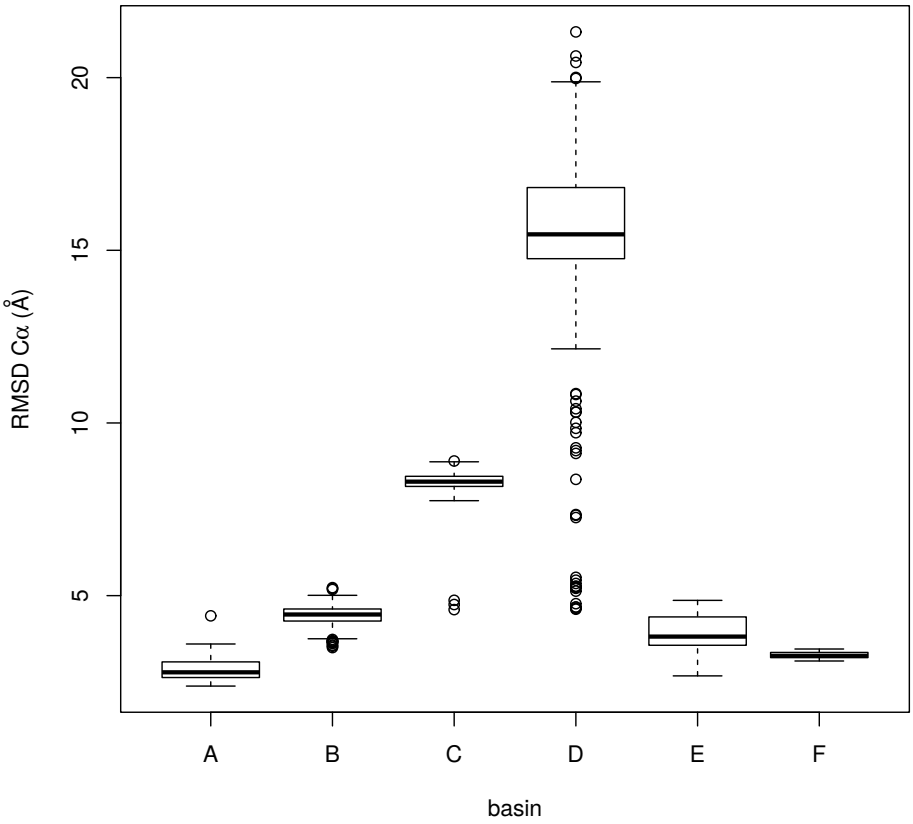


Figure 4.17: The C_α root-mean-square deviation (RMSD), relative to the native state, for each conformation in the energy basins is shown as a box and whisker plot.

4.5 Material and Methods

4.5.1 TDP-43 RRM2 structure preparation.

The structure of RRM2 of TDP-43 used in the simulation, residues 189 to 261, was obtained from the deposited NMR solution structure (pdb entry: 1WF0). Wild-type mutations S191R, G192K, and G200E were introduced in the sequence of RRM2 using the Mutator plugin (v. 1.3) of VMD [194]. The resulting structure was solvated using GROMACS [267, 268, 269, 270] with 37000 water molecules in a dodecahedron water box (1140 nm^3), containing 5 Na^+ ions to enforce charge neutrality.

4.5.2 NMR assignment of TDP-43 RRM2.

Backbone chemical shifts of TDP-43 RRM2 at 6 M concentration of urea were collected as previously described by Mackness et al. [252] (BMRB deposition number 27549).

4.5.3 Simulation protocol.

Molecular dynamics simulations of the RRM2 of TDP-43 were performed in GROMACS using the Amber03W force field [271] with the TIP4P05 water model [272]. The equations of motion were integrated using the LINCS constraints to allow a time step of 2 fs [273]. The van der Waals interactions were implemented with a cutoff distance of 0.9 nm and the particle mesh Ewald method was used to treat electrostatic interactions with periodic boundary conditions [274]. All simulations were carried out in the canonical ensemble at constant volume, and temperature was maintained with the Nosé-Hoover thermostat [275].

A 30 ns unfolding trajectory at 450 K was preliminarily collected, starting from the solvated structure described above, to obtain four starting conformations for the replica-averaged metadynamics (RAM) simulations [250, 251]. The C_α root-mean-square deviation (RMSD) of the resulting structures varies between 2.5 to 3 Å. Each conformation was subsequently relaxed at 300 K for 10 ns. PLUMED2 [276] and

GROMACS were used to perform the RAM simulations using chemical shifts as replica-averaged restraints [251, 277] and bias-exchange metadynamics [250]. Four replicas of the system were run in parallel at 300 K with restraints from experimental data applied at each time step to the average values of the NMR chemical shifts calculated with the Camshift algorithm [278, 279]:

$$E^{cs} = \alpha \sum_{k=1}^{N} \sum_{l=1}^{4} \left(\delta_{kl}^{exp} - \frac{1}{M} \sum_{m=1}^{M} \delta_{klm}^{calc} \right)^2$$

where the α is the force constant, set to 12 kJ/(mol·ppm^2), k runs over the amino acids of the protein, l runs over the four backbone atoms used in the simulatiions (Cα, Cβ, H_N and N) and m runs over the $M = 4$ replicas. The chemical shifts of proline, glycine, aspartate, glutamate and histidine residues were not included in the simulations [71]. In this way the system evolves with a force field that is perturbed in such a way to increase the agreement with the experimental chemical shifts as resulting by the application of the maximum entropy principle [71]. The experimental chemical shifts were collected at 6 M urea and at this concentration of urea the intermediate state is in fast exchange with the unfolded state. Thus, the measured chemical shifts are the population weighted average of the 2 states [252]. The population of the unfolded state is estimated to be less than 5% [252]. In a previous study, Mielke et al. [255] have shown that reference random coil chemical shifts collected in 8 M urea for ^{13}Cα and ^{13}Cβ have modest to negligible differences to those collected in aqueous solution. Therefore, no correction are required upon use of chemical shifts collected at 6 M urea in the metadynamics simulations.

Each replica of the system employed a metadynamics approach to bias a collective variable (CV), with exchanges between the replicas attempted every 50 ps according to a replica-exchange scheme [65]. The CVs used in the simulations were calculated as described:

- Total α-helical content (αC): the ALPHARMSD module of PLUMED2 was used, with a rational switching function [280] ($r_0 = 0.08$ nm, $n = 8$, $m = 12$).

- Total β-sheet content (βC): the ANTIBETARMSD module of PLUMED2 was used, with a rational switching function [280] ($r_0 = 0.08$ nm, $n = 8$, $m = 12$).

- Radius of gyration (R_g): the GYRATION module of PLUMED2 was used.

- Number of hydrophobic contacts within the ILV cluster (#ILV): the CONTACTMAP module of PLUMED2 was used. The presence of each of the eighteen contacts identified between isoleucine, leucine and valine residues for the native state of the RRM2 of TDP-43 [236, 143] was evaluated using a rational switching function with $r_0 = 0.6$ nm for the distance between the Cβ atoms of each pair of residues.

Gaussians deposition was performed with an initial rate of 0.125 kJ/mol/ps, where the σ values were set to 0.46, 0.32, 0.006, and 0.34 nm, for αC, βC, R_g and #ILV, respectively. The σ values were obtained as the average fluctuations of each CV from unbiased simulations.

4.5.4 RAM simulations analysis.

The RAM simulations were analyzed using the METAGUI plugin of VMD [281]. Configurations corresponding to the partially structured states of RRM2 were grouped in microstates by dividing the 4-dimensional CV-space in hypercubes [281]. The free energy of the microstates was then computed with the WHAM technique [282, 283]. These microstates were clustered [284] to identify the energy basins in the free energy landscape of RRM2 (Fig. 4.8). For each conformation the difference from the structure of the native state was calculated and shown as RMSD of the C_α in Fig. 4.17.

As previously described [71]. the convergence of the sampling was assessed by monitoring the differences of the free energies at increasing simulation length during

the simulations. After the first 570 ns the free energy landscapes for each CV were stable below 2 kJ/mol (Fig. 4.6), thus the sampling was converged.

Hydrophobic ILV contacts for the structures resulting from the simulations were estimated as the amount of surface area buried between a pair of residues: $SASA_a$+$SASA_b$-$SASA_{a+b}$, with a and b being ILV residues and SASA being the solvent accessible surface area of the side chain of residue a, b, and residues a and b together. The presence of an ILV contact was computed, using VMD, as 1 if two residues have at least 45 $Å^2$ of buried surface area and 0 otherwise [236]. The probability of an ILV contact was calculated as the occurrence of the contact in the microstates of each basin.

Aggrescan3D [259] calculations were performed using the Aggrescan3D server with a distance of aggregation analysis of 5 Å. Twentyfive structures for each basin were uploaded to the server and the average value and standard deviation of each residue were used for graphical representation.

CHAPTER 5

DISCUSSION

In the past decades, the idea that proteins might employ disordered domains to accomplish their biological function has been confirmed by an increasing number of experimental studies [285]. In particular, disordered domains play a major role in protein-protein and protein-nucleic acid interactions: the lack of structural constraints, in fact, allows proteins to adapt their conformation to multiple targets; moreover, the equilibrium between folded and unfolded states may vary in response to external factors, tuning the activity of disordered proteins to respond to different stimuli. Many disordered proteins are involved in fundamental processes in the organism, as in the case of transcription factors and cytokines regulators. Consequently, dysregulation of the activity of these proteins, caused by mutations or external factors, often correlates with pathogenic conditions. By obtaining insights into the conformations sampled in disordered states, it is possible to elucidate the mechanisms that govern normal and aberrant protein functions and possibly to identify therapeutic targets. The main challenge in studying these systems, however, consists in their intrinsic structural dishomogeneity that results in the experimental inaccessibility of these disordered states with the traditional techniques of structural biology [59]. Therefore, new experimental approaches are necessary to explore the energy landscape of these proteins.

In this thesis work, I have elucidated how structural disorder contributes to the activity of three RNA-binding proteins. In particular, I focused on the human proteins TTP and TDP-43 and the *C. elegans* protein MEX-5. TTP and MEX-5 are mem-

bers of the same CCCH-type tandem zinc finger protein family, and both contain an unfolded zinc finger in their RNA-free form, ZF2 for TTP and ZF1 for MEX-5, respectively. A combination of computational and experimental techniques was employed to identify the intramolecular interactions that determine the structural differences between the folded and unfolded zinc fingers in each domain. Altogether, the results presented in this thesis demonstrate that these two proteins have evolved, independently, to modulate their activity through their structure.

Additionally, the structural characterization of the partially structured folding intermediate of TDP-43 RRM2, described in this thesis, contribute to elucidate whether this state may have a functional purpose, for example in cytoplasmic transport, or may mediate misfolding and promote aggregation. The insights presented in this work may drive design of specific antibodies for therapeutic treatments of neurodegeneration.

5.1 Major results and implications

5.1.1 A π-π stacking interaction within the zinc fingers of Tristetraprolin and TIS11d stabilizes the coordination of a zinc ion

In Chapter 2, I discovered which elements of the primary sequence, in the TIS11 family of proteins, determine the folded/unfolded character of the C-terminal zinc finger. TTP and TIS11d, two members of the human CCCH-type tandem zinc finger proteins, share large sequence identity. Despite this similarity, and a similar specificity and affinity towards the same RNA targets, the N-terminal zinc finger (ZF1) of TTP is unstructured and does not coordinate zinc in the RNA-free form, but folds upon RNA-binding. In TIS11d, however, both the zinc fingers (ZF1 and ZF2) are folded in the RNA-free form. Recent results from our lab [96] have shown, using a reporter luciferase assay in cell, that the mRNA destabilizing activity of TTP was increased when the partially disordered RNA-binding domain of TTP was replaced with the

fully structured domain of TIS11d. These results demonstrate that the degree of structure of the RNA-binding domain directly affects the function of the protein in the cell, suggesting that protein activity is modulated through its structure. Deveau et al. [96] have also demonstrated that the residues in the C-terminal part of the short α-helix, within the first two zinc-ligand cysteines, determines the folded/unfolded character of the zinc finger. However, the interactions involved, and the mechanism through which these residues regulate the zinc affinity, were not determined.

The unfolded nature of the C-terminal zinc finger of TTP represented an obstacle for structural characterization of the interactions that stabilize the fold of the domain. Using molecular dynamics simulations and homology modeling, however, I was able to overcome this problem. Exploiting the *in silico* model of the folded C-terminal zinc finger of TTP that is coordinating zinc with both zinc fingers, I have been able to follow the series of events that lead to loss of zinc coordination and unfolding, highlighting the key interactions involved. Explicitly, I found that the imidazole ring of the zinc-ligand histidine is restrained to a rotameric state compatible with zinc binding by a stacking interaction with a conserved aromatic side chain. In both TIS11d zinc fingers, the stacking between the aromatic moiety and the histidine is stabilized by a hydrogen bond that occurs within the short α-helix between the first two ligand cysteines. In TTP, however, such hydrogen bond only occurs in the N-terminal zinc finger. Interestingly, the residue involved in the hydrogen bond, in TIS11d ZF2, is the threonine within the HTI stabilizing motif identified by Deveau et al. [96]. The corresponding motif in TTP ZF2, YLQ, contains a leucine, whose side chain cannot form hydrogen bond. Therefore, the results presented in Chapter 2 are in agreement with previous studies, and, furthermore, they provide insights into the molecular mechanism by which zinc coordination is stabilized in TTP and TIS11d.

Because CCCH-type zinc fingers are conserved motifs in proteins across many species, spanning from yeast to human, we sought to investigate if regulation of

activity through the structure, or lack of, in the RNA-binding domain is a common mechanism in this class of proteins. The sequence alignment of 14,851 CCCH-type zinc fingers reveals that the aromatic character of the third residue following the first ligand cysteine is highly conserved ($> 94\%$), indicating that the stacking interaction described for TTP and TIS11d is likely present in most CCCH-type fingers. The threonine, or serine, in the stabilizing HTI motif, however, is present in roughly half of the analyzed sequences. Based on our results, I hypothesize that the remaining zinc fingers might have unstructured domains as TTP; alternatively, they might use alternative mechanisms to stabilize the coordination of zinc. In order to determine if other proteins exploit their thermodynamic stability to modulate their activity, the methods described in Chapter 2, and applied to TTP, can be used to characterize other CCCH-type fingers.

5.1.2 The unfolded state of MEX-5 ZF1 is likely relevant to *Caenorhabditis elegans* reproductive physiology at the oocyte-to-embryo transition

In Chapter 3, I demonstrated that the CCCH-type tandem zinc finger protein MEX-5, of the nematode *Caenorhabditis elegans*, contains an unstructured zinc finger in the RNA-free state, but coordinates zinc with both zinc fingers in its complexed form with RNA. Moreover, I showed that this unfolded zinc finger is required for the physiological function of the protein. MEX-5 is an RNA-binding protein that plays a crucial role in early development of *C. elegans*. During embryogenesis, in fact, MEX-5 distributes in the zygote cytoplasm along an anterior-high posterior-low concentration gradient [111], that in turn governs the segregation in the posterior of several cell fate determinant, such as POS-1 and PIE-1. Additionally, recent studies [116, 117, 118] have shown that MEX-5 regulates the formation of germline granules. Although several elements of the primary sequence of the tandem zinc finger domain of MEX-5

diverge from other members of the protein family, their contribution to MEX-5 activity has not been explored. Therefore, in order to elucidate the relationship between primary sequence and protein activity, I proceeded to characterize the structure and dynamics of the tandem zinc finger domain of MEX-5.

Using nuclear magnetic resonance spectroscopy, and in particular heteronuclear single quantum coherence experiments, I demonstrated that the tandem zinc finger domain of MEX-5 is partially unstructured in the RNA-free form. I showed that the N-terminal zinc finger, in fact, only folds and coordinates zinc upon binding with RNA, unlike the C-terminal zinc finger that stably coordinates zinc in both the free and the bound forms. Although this mechanism is reminiscent of the human homolog TTP, I showed that the interactions that define the stability of the zinc fingers diverge from the other members of the CCCH-type zinc fingers proteins. The solution structure of the C-terminal zinc finger of MEX-5 presented in Chapter 3, in fact, reveals that the backbone architecture of the finger diverges from the human homologs TTP and TIS11d. Specifically, the region between the first two ligand cysteines contains only a short turn of 3_{10}-helix, in place of the α-helix of the human proteins, and a flexible glycine-rich loop. As described in Chapter 2, the conformation of this region in TTP and TIS11d is crucial for posing a conserved aromatic side chain to stack against the ligand histidine, thus stabilizing the zinc coordination and the fold of the finger. For MEX-5, molecular dynamics simulations confirmed that the stacking interaction occurs in the C-terminal zinc finger but not in the N-terminal one, recapitulating our hypothesis, formulated in Chapter 2, that the π-π interaction is crucial for the stability of the finger structure. Based on these results, I have been able to design a variant of MEX-5 in which both zinc fingers are folded in the RNA-free form. This variant of MEX-5 was as an invaluable tool to investigate the role of the disorder-to-order transition of the tandem zinc finger domain of MEX-5, upon binding to RNA, to the protein physiological activity. In collaboration with Sean Ryder's group, we employed

a CRISPR/Cas9 homologous recombination approach to introduce the MEX-5 variant in the endogenous locus of *C. elegans*. The resulting animals, homozygous for the mutation, exhibit sterility, embryonic lethality and develop uterine embryoid-like bodies, features that do not correspond with the *null* phenotype [108]. Therefore, we concluded that the unfolded state of the N-terminal zinc finger of MEX-5 in the RNA-free state is relevant for the reproductive physiology of *C. elegans*, through a mechanism that has not been elucidated yet.

The results presented in Chapter 3 open new perspectives for understanding the role of MEX-5 during embryogenesis. Whether or not the phenotype observed in the *mex-5(spr1)* worms is related to an alteration of known MEX-5 functions, such as its spatial distribution pattern [111] or the recruiting of PLK-1 [115], is still to be determined. However, we showed that the disorder-to-order transition upon RNA-binding of a zinc finger is a mechanism that appears conserved across species in the CCCH-type tandem zinc finger family. Interestingly, the interactions that define this transition are not conserved. We showed that the architecture of MEX-5 fingers differs from the otherwise invariant structure of other CCCH-type proteins [172, 90][Ertekin A., Massi F., unpublished data]. This observation suggests that certain CCCH-type zinc finger proteins have evolved independently to exploit their thermodynamic stability to regulate their biological functions, as in the case of MEX-5 and TTP.

5.1.3 Partially folded state of TDP-43 RRM2 expose to the solvent the aggregation-prone peptide regions located on strands β3, β4 and β5

In Chapter 4, I proposed a potential pathogenic role in ALS for the folding intermediate state previously identified by Mackness et al. [143]. TDP-43 is a major component of the aggregates that are the hallmark of ALS and other neurodegenerative diseases. In TDP-43, both the low complexity C-terminal domain and the second RNA-recognition motif (RRM2) have been shown to have an intrinsic aggre-

gation and amyloidogenic propensity [122, 159]. The structural changes that might facilitate TDP-43 aggregation, however, are still unclear. Mackness et al. provided insights into this mechanism when they characterized the folding pathway of TDP-43 RRM2, discovering that it could populate at least one intermediate state [143]. Although several structural features of the partially folded intermediate have been unveiled by Mackness et al. [Mackness B.C., Morgan B.R., Deveau L.M., Kathuria S.V., Tavella D., Massi F., Zitzewitz J.A., unpublished data], a description at atomic resolution of this state, that is required in order to establish its role in misfolding and aggregation, was not available. This state, in fact, is rarely populated under native conditions and it is characterized by an high structural heterogeneity, two factors that complicate the use of traditional methods of structural biology.

Using an advanced computational approach involving replica averaged metadynamics, alongside experimental NMR chemical shifts collected at 6 M urea denaturant, I was able to reveal the structural details of the partially folded states of TDP-43 RRM2 that correspond to the folding intermediate. These states display an overall progressive degree of unfolding, corresponding to loss of secondary structure elements and hydrophobic contacts. The most unfolded states, however, exhibit an increase in helical structure, with a non-native α helix occurring in place of the strand β_2. I observed that a common feature of these states is the exposure to the solvent of aggregation-prone residues in strands β_3,β_4 and β_5, suggesting that the unfolding of RRM2 may play an important role in the conversion of TDP-43 into the pathological aggregates that occur in neurodegenerative diseases. Moreover, the partially folded states may contribute to the nucleocytoplasmic trafficking of TDP-43 by a similar mechanism, increasing the exposure of the nuclear export signal present in RRM2. However, whether or not the exposure of the NES sequence of TDP-43 in the intermediate state plays an active role in cytoplasmic transport has yet to be determined.

The structural characterization of the partially folded states of TDP-43 RRM2,

presented in Chapter 4, provides the basis for further studies that will explore protein misfolding and amyloid formation in TDP-43 proteinopathies and other neurodegenerative diseases. The relevance of the partially folded states of RRM2 in pathogenesis, in fact, is supported by previous works [286], showing that other RNA-binding domains in disease-related proteins, FUS/TLS, matrin-3 and hnRNP A1, also populate a stable intermediate state. A key gap in knowledge to how these partially folded states contributes to physiological functions (RNA binding, nuclear trafficking) and/or to pathogenic truncation, aggregation and nuclear depletion. The results described in Chapter 4 provide a link between these partially unfolded states and their increased exposure of aggregation-prone residues, thus supporting their potential relationship to pathogenic mechanisms of TDP-43-associated neurodegeneration. Further experimental approaches including mutagenesis, however, are required to validate this hypothesis both *in vitro* and *in vivo* ALS models.

5.2 Future directions

In this thesis I have explored the role of structure-disorder conversion and intrinsic disorder in the functions of three RNA-binding proteins. In particular, I focused on the central problem of developing an appropriate set of experimental and computational techniques to capture, at atomic resolution, the structures and dynamics of IDPs and IDRs. The characterization of the heterogeneous structural ensemble of disordered or partially disordered domains, in fact, represents the main challenge in the study of IDPs. The relevance of the approaches described in this work extends beyond the analysis of the studied systems and these methods are suitable to be applied to the study of other proteins that contain disordered regions in their native state.

Molecular dynamics simulations have been extensively employed for studying protein folding and the underlying free-energy landscapes. In the particular case of IDPs,

unstructured states are often experimentally inaccessible or only collection of sparse data is possible. Molecular dynamics simulations, with advanced methods such as enhanced sampling, represent a way to explore regions of the free-energy landscape of these systems beyond the native state. Moreover, traditional force fields used in MD simulations can be extended to include experimental data in order to provide a better representation of the structural ensemble in the disordered state.

In the studies presented in this thesis, proper description of the interactions between the zinc ion and the ligand residues, through incorporation of charge transfer and polarization effects terms in the CHARMM force field, allowed to identify the key interactions that determine the stability of the native conformation in tandem zinc finger proteins. The approach applied here to the tandem zinc finger domains of TTP and MEX-5 may be extended to other zinc-binding protein that display a disorder-to-order transition. In particular, the combination of homology modeling and MD simulations might be beneficial to provide structural insights into these disordered domains and drive the identification of mutations that alter the order/disorder equilibrium.

In addition, the results presented in Chapter 2 and 3 represent, to our knowledge, one of the few example of how the unstructured states of proteins that undergo to a disorder to order transition upon binding might be functional. Previous studies have explored the role of disordered states on the biological activity of proteins [220, 96], however here we provided a structural characterization of these states and we proved that the extent of structure in the disordered state affects the biological function of the protein.

In Chapter 4, we showed that structural disorder can be functional for the physiological or pathological functions of a protein even if unstructured states are populated at low percentage under native condition. Folding intermediates, in fact, contains unstructured regions in a similar fashion to IDPs. However, because of their transient

nature, experimental characterization of folding intermediates might be challenging. The methods used to characterize the free-energy landscape of TDP-43 RRM2 might be extended to other proteins whose folding pathway includes one or more functionally significant intermediate states. The RAM approach, in fact, allows inclusions of a diverse set of experimental data and a proper choice of collective variable ensures an effective sampling of the conformational space. An alternative approach to study these transient states is offered by relaxation dispersion nuclear magnetic resonance experiments [287, 288]. This NMR method is used to quantify millisecond time scale dynamics that involve transitions between different protein conformations. Data-fitting procedures allow then to extract the kinetic and the thermodynamic properties of the exchange process and the structural features of the invisible excited states along the exchange pathway. The relaxation dispersion approach facilitates detection of intermediates and other excited states that are populated at low levels, 0.5% or higher, that cannot be observed directly in NMR spectra, so long as they exchange with the observable ground state of the protein on the millisecond time scale.

5.2.1 Understanding the molecular basis of leukemia-associated mutations in TIS11 proteins

The human TIS11 proteins, including TTP and TIS11d, recognize the AU-rich elements in the 3′-UTR of several mRNAs, and thus promote their turnover. TTP and TIS11d regulate the expression of genes encoding inflammation- and cancer-associated factors and, therefore, dysregulation of TIS11 proteins function is associated with cancer, chronic inflammation and auto-immune pathologies.

Several mutations in the RNA-binding domain of TIS11d have been identified in patients with acute myeloid leukemia, acute lymphoblastic leukemia and in lymphoma cell lines [99]. Previous studies [100] have shown that the effect of mutations P190L

and D219E in TIS11d is to destabilize the fold of the C-terminal zinc finger in the RNA-free state, making the domain more similar to TTP. Because the decreased degree of structure in the TFZ domain in TIS11 proteins has been demonstrated to cause reduced mRNA turnover in the cell [96], the effect of this mutations is likely to decrease TIS11d activity and, thus, to upregulate its targets.

In order to mitigate the effect of these pathological mutations, we can leverage our knowledge of the stabilizing interactions identified in Chapter 2 for rational design of small molecules. The short α-helix between the first two ligand cysteines, whose conformation is critical for zinc coordination, represents a good candidate for small molecule targeting. α-helices, in fact, are widely used as templates for inhibitor design, and there is a vast literature describing small non-peptide molecules that bind to an α-helical peptide receptor [289]. The effect of the candidate small molecules in stabilizing the fold of the zinc finger can be evaluated *in silico*, with the molecular dynamics simulations described in Chapter 2, and *in vitro*, using ^{15}N-^{1}H HSQC.

5.2.2 Investigating the role of the unfolded state of MEX-5 ZF1 during embryogenesis

The results presented in Chapter 3 demonstrated that the unfolded state of ZF1 is required for the physiological function of MEX-5 during embryogenesis. However, because the strain *mex-5(spr1)* has been obtained through a CRISPR/Cas9 approach, the observed phenotype might be due to off-target RNA-guided endonuclease-induced mutations at sites other than the intended on-target site [290]. To rule out this possibility, whole genome sequencing will be performed on the *mex-5(spr1)* worms.

Additionally, in order to determine which aspect of MEX-5 function is affected by the unfolded/folded state of ZF1, the cytoplasmic distribution of the $CX_{10}C$ mutant MEX-5 in the zygote needs to be evaluated. One possible approach is to generate transgenic lines expressing GFP::MEX-5$_{CX_{10}C}$ or GFP::MEX-5 wild type [111], and

to monitor the protein localization using live fluorescence microscopy.

Moreover, additional studies are needed to determine at which stage of embryogenesis the folded state of ZF1 in MEX-$5_{CX_{10}C}$ diverges from the wild type protein function. To this aim, a RNAi screening targeting genes involved in cell division processes or cell maintenance processes [291] might be a valuable tool.

5.2.3 Validating the pathogenic role of the partially folded states of TDP-43 RRM2

To expand our knowledge on the mechanisms that lead to TDP-43 pathological aggregation in ALS and other TDP-43 proteinopathies, and in particular to validate the hypothesis that that partial or complete unfolding of the RRM2 may contribute to TDP-43 misfolding and amyloid aggregates, additional experiments, including mutagenesis, are necessary. In Chapter 3, we propose that mutations of Val 193, Val 195 and Val 232 that decrease the hydrophobic moiety of the residues (i.e., V193A or V195A or V232A), will destabilize the most native-like states in favor of the less structured and more extended states. We predict that the aggregation propensity of these mutant TPD-43 variants will be increased compared to the wild type and similar to the truncated C-terminal fragments.

To date, there are no biomarkers that can track TDP-43 misfolding and aggregation, and no TDP-43-based therapies are available for any of the numerous diseases associated with it. The structural insights into the partially folded states of TDP-43 RRM2, provided in Chapter 4, could inform the design of therapeutics, such as antibodies recognizing specific elements of secondary structures or small molecule compounds that inhibit the initial formation of these predicted toxic conformations. Explicitly, I propose the non-native α-helix, that occurs in place of the strand β_2 and is observed in the least-structured states of RRM2, as a candidate for antibody targeting, to improve both the diagnosis and the therapy of TDP-43 proteinopathies.

5.3 Concluding remarks

Many proteins, or regions of proteins, are intrinsically disordered under native, functional conditions. Intrinsically disordered proteins present a challenge for structure determination, by experiments alone, because they populate highly heterogeneous ensembles of conformations. Moreover, the inherent dynamics of these unstructured proteins hamper detailed analysis of their structure/function relationships. Computational methods currently offer the best opportunity to study the conformations of these proteins, and they also enable the mapping of the free energy landscape of the structural ensemble.

The results presented in these thesis show that computational studies and experimental techniques, used in concert, are a powerful tool to learn about the structure and dynamics of disordered proteins in physiological condition. In this work, I showed that the balance between order and disorder in TTP, MEX-5 and TDP-43 is essential for their function. In particular, I highlighted the importance of characterizing and identifying the elements that affect this balance to fully understand the function of a protein, and how disrupting this balance could lead to disease.

I hope my work on the structural characterization of disordered proteins will serve as a helpful guide for further investigation of the balance between order and disorder, and might provide insights on disease mechanisms and inform therapeutic development.

BIBLIOGRAPHY

[1] Fischer, Emil. Einfluss der configuration auf die wirkung der enzyme. *Berichte der deutschen chemischen Gesellschaft 27*, 3 (1894), 2985–2993.

[2] Wu, Hsien. Studies on denaturation of proteins xiii. a theory of denaturation. In *Advances in protein chemistry*, vol. 46. Elsevier, 1995, pp. 6–26.

[3] Mirsky, Alfred E, and Pauling, Linus. On the structure of native, denatured, and coagulated proteins. *Proceedings of the National Academy of Sciences of the United States of America 22*, 7 (1936), 439.

[4] Uversky, Vladimir N. Natively unfolded proteins: a point where biology waits for physics. *Protein science 11*, 4 (2002), 739–756.

[5] Dunker, A Keith, Lawson, J David, Brown, Celeste J, Williams, Ryan M, Romero, Pedro, Oh, Jeong S, Oldfield, Christopher J, Campen, Andrew M, Ratliff, Catherine M, Hipps, Kerry W, et al. Intrinsically disordered protein. *Journal of molecular graphics and modelling 19*, 1 (2001), 26–59.

[6] Dyson, H Jane, and Wright, Peter. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Advances in protein chemistry 62* (2002), 311–340.

[7] Boesch, Chris, Bundi, Arno, Oppliger, Max, and Wüthrich, Kurt. 1h nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *European journal of biochemistry 91*, 1 (1978), 209–214.

[8] Daniels, AJ, Williams, RJP, and Wright, PE. The character of the stored molecules in chromaffin granules of the adrenal medulla: a nuclear magnetic resonance study. *Neuroscience 3*, 6 (1978), 573–585.

[9] Qian, Yan Qiu, Otting, Gottfried, Furukubo-Tokunaga, Katsuo, Affolter, Markus, Gehring, Walter J, and Wüthrich, K. Nmr structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the drosophila antennapedia protein. *Proceedings of the National Academy of Sciences 89*, 22 (1992), 10738–10742.

[10] Trombitás, Karoly, Greaser, Marion, Labeit, Siegfried, Jin, Jian-Ping, Kellermayer, Miklós, Helmes, Michiel, and Granzier, Henk. Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. *The Journal of cell biology 140*, 4 (1998), 853–859.

[11] Rauscher, Sarah, and Pomès, Régis. Structural disorder and protein elasticity. In *Fuzziness*. Springer, 2012, pp. 159–183.

[12] Hoh, Jan H. Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. *Proteins: Structure, Function, and Bioinformatics 32*, 2 (1998), 223–228.

[13] Uversky, Vladimir N, Narizhneva, Natalya V, Ivanova, Tatyana V, Kirkitadze, Marina D, and Tomashevski, Andrey Yu. Ligand-free form of human α-fetoprotein: evidence for the molten globule state. *FEBS letters 410*, 2-3 (1997), 280–284.

[14] Ebert, Marc-Olivier, Bae, Sung-Hun, Dyson, H Jane, and Wright, Peter E. Nmr relaxation study of the complex formed between cbp and the activation domain of the nuclear hormone receptor coactivator actr. *Biochemistry 47*, 5 (2008), 1299–1308.

[15] Dyson, H Jane, and Wright, Peter E. Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology 6*, 3 (2005), 197.

[16] Wright, Peter E, and Dyson, H Jane. Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews Molecular cell biology 16*, 1 (2015), 18.

[17] Wright, Peter E, and Dyson, H Jane. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *Journal of molecular biology 293*, 2 (1999), 321–331.

[18] Oldfield, Christopher J, Cheng, Yugong, Cortese, Marc S, Romero, Pedro, Uversky, Vladimir N, and Dunker, A Keith. Coupled folding and binding with α-helix-forming molecular recognition elements. *Biochemistry 44*, 37 (2005), 12454–12470.

[19] Pontius, Brian W. Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends in biochemical sciences 18*, 5 (1993), 181–186.

[20] Gsponer, Jörg, and Babu, M Madan. The rules of disorder or why disorder rules. *Progress in biophysics and molecular biology 99*, 2-3 (2009), 94–103.

[21] Lee, Chul Won, Ferreon, Josephine C, Ferreon, Allan Chris M, Arai, Munehito, and Wright, Peter E. Graded enhancement of p53 binding to creb-binding protein (cbp) by multisite phosphorylation. *Proceedings of the National Academy of Sciences* (2010).

[22] Borg, Mikael, Mittag, Tanja, Pawson, Tony, Tyers, Mike, Forman-Kay, Julie D, and Chan, Hue Sun. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proceedings of the National Academy of Sciences 104*, 23 (2007), 9650–9655.

[23] Van Roey, Kim, Dinkel, Holger, Weatheritt, Robert J, Gibson, Toby J, and Davey, Norman E. The switches. elm resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal. 6*, 269 (2013), rs7–rs7.

[24] Van Roey, Kim, Gibson, Toby J, and Davey, Norman E. Motif switches: decision-making in cell regulation. *Current opinion in structural biology 22*, 3 (2012), 378–385.

[25] Dunker, A Keith, and Obradovic, Zoran. The protein trinitylinking function and disorder. *Nature biotechnology 19*, 9 (2001), 805.

[26] Holmes, KC. Flexibility in tobacco mosaic virus. *Mobility and Function in Proteins and Nucleic Acids* (1983), 116.

[27] Reichmann, Dana, and Jakob, Ursula. The roles of conditional disorder in redox proteins. *Current opinion in structural biology 23*, 3 (2013), 436–442.

[28] Fuxreiter, Monika, Simon, Istvan, Friedrich, Peter, and Tompa, Peter. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of molecular biology 338*, 5 (2004), 1015–1026.

[29] Burgen, ASV, Roberts, GCK, and Feeney, J. Binding of flexible ligands to macromolecules. *Nature 253*, 5494 (1975), 753.

[30] Koshland Jr, DE. Enzyme flexibility and enzyme action. *Journal of cellular and comparative physiology 54*, S1 (1959), 245–258.

[31] Spolar, Ruth S, and Record, M Thomas. Coupling of local folding to site-specific binding of proteins to dna. *Science 263*, 5148 (1994), 777–784.

[32] Abate, Cory, Patel, Lekha, Rauscher, FJ, 3rd, and Curran, Tom. Redox regulation of fos and jun dna-binding activity in vitro. *Science 249*, 4973 (1990), 1157–1161.

[33] Laity, John H, Dyson, H Jane, and Wright, Peter E. Dna-induced α-helix capping in conserved linker sequences is a determinant of binding affinity in cys2-his2 zinc fingers1. *Journal of molecular biology 295*, 4 (2000), 719–727.

[34] DiNitto, Jonathan P, and Huber, Paul W. Mutual induced fit binding of xenopus ribosomal protein l5 to 5 s rrna. *Journal of molecular biology 330*, 5 (2003), 979–992.

[35] Rosenbaum, Joel C, Fredrickson, Eric K, Oeser, Michelle L, Garrett-Engele, Carrie M, Locke, Melissa N, Richardson, Lauren A, Nelson, Zara W, Hetrick, Elizabeth D, Milac, Thomas I, Gottschling, Daniel E, et al. Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates. *Molecular cell 41*, 1 (2011), 93–106.

[36] Das, Rahul K, and Pappu, Rohit V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences 110*, 33 (2013), 13392–13397.

[37] Venkatraman, Prasanna, Wetzel, Ronald, Tanaka, Motomasa, Nukina, Nobuyuki, and Goldberg, Alfred L. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Molecular cell 14*, 1 (2004), 95–104.

[38] McMeekin, TL, and Marshall, Kathleen. Specific volumes of proteins and the relationship to their amino acid contents. *Science 116*, 3006 (1952), 142–143.

[39] Jirgensons, B. Optical rotation and viscosity of native and denatured proteins. xi. relationships between rotatory dispersion, ionization and configuration. *Archives of biochemistry and biophysics 74*, 1 (1958), 70–83.

[40] Doolittle, RF. Structural aspects of the fibrinogen to fibrin conversion. In *Advances in protein chemistry*, vol. 27. Elsevier, 1973, pp. 1–109.

[41] Manalan, Allan S, and Klee, Claude B. Activation of calcineurin by limited proteolysis. *Proceedings of the National Academy of Sciences 80*, 14 (1983), 4291–4295.

[42] De Kruif, Cornelis G, Huppertz, Thom, Urban, Volker S, and Petukhov, Andrei V. Casein micelles and their internal structure. *Advances in Colloid and Interface Science 171* (2012), 36–52.

[43] Byrne, B Marion, Van het Schip, Alfred D, Van de Klundert, Jos AM, Arnberg, Annika C, Gruber, Max, and Ab, Geert. Amino acid sequence of phosvitin derived from the nucleotide sequence of part of the chicken vitellogenin gene. *Biochemistry 23*, 19 (1984), 4275–4279.

[44] Tsurupa, Galina, Mahid, Ariza, Veklich, Yuri, Weisel, John W, and Medved, Leonid. Structure, stability, and interaction of fibrin αc-domain polymers. *Biochemistry 50*, 37 (2011), 8028–8037.

[45] Klee, CB, Crouch, TH, and Krinks, MH. Calcineurin: a calcium-and calmodulin-binding protein of the nervous system. *Proceedings of the National Academy of Sciences 76*, 12 (1979), 6270–6273.

[46] Stewart, Alexander A, Ingebritsen, Thomas S, Manalan, Allan, Klee, Claude B, and Cohen, Philip. Discovery of a ca2+-and calmodulin-dependent protein phosphatase. *FEBS letters 137*, 1 (1982), 80–84.

[47] Kissinger, Charles R, Parge, Hans E, Knighton, Daniel R, Lewis, Cristina T, Pelletier, Laura A, Tempczyk, Anna, Kalish, Vincent J, Tucker, Kathleen D, Showalter, Richard E, Moomaw, Ellen W, et al. Crystal structures of human calcineurin and the human fkbp12–fk506–calcineurin complex. *Nature 378*, 6557 (1995), 641.

[48] Rumi-Masante, Julie, Rusinga, Farai I, Lester, Terrence E, Dunlap, Tori B, Williams, Todd D, Dunker, A Keith, Weis, David D, and Creamer, Trevor P. Structural basis for activation of calcineurin by calmodulin. *Journal of molecular biology 415*, 2 (2012), 307–317.

[49] Arany, Zoltàn, Huang, L Eric, Eckner, Richard, Bhattacharya, Shoumo, Jiang, Chian, Goldberg, Mark A, Bunn, H Franklin, and Livingston, David M. An essential role for p300/cbp in the cellular response to hypoxia. *Proceedings of the National Academy of Sciences 93*, 23 (1996), 12969–12973.

[50] Lando, David, Peet, Daniel J, Whelan, Dean A, Gorman, Jeffrey J, and Whitelaw, Murray L. Asparagine hydroxylation of the hif transactivation domain: a hypoxic switch. *Science 295*, 5556 (2002), 858–861.

[51] Elkins, Jonathan M, Hewitson, Kirsty S, McNeill, Luke A, Seibel, Jürgen F, Schlemminger, Imre, Pugh, Christopher W, Ratcliffe, Peter J, and Schofield, Christopher J. Structure of factor-inhibiting hypoxia-inducible factor (hif) reveals mechanism of oxidative modification of hif-1α. *Journal of Biological Chemistry 278*, 3 (2003), 1802–1806.

[52] Dann, Charles E, Bruick, Richard K, and Deisenhofer, Johann. Structure of factor-inhibiting hypoxia-inducible factor 1: an asparaginyl hydroxylase involved in the hypoxic response pathway. *Proceedings of the National Academy of Sciences 99*, 24 (2002), 15351–15356.

[53] Schulz, GEORG E. Nucleotide binding proteins. *Molecular mechanism of biological recognition* (1979), 79–94.

[54] Shammas, SL, Rogers, JM, Hill, SA, and Clarke, J. Slow, reversible, coupled folding and binding of the spectrin tetramerization domain. *Biophysical journal 103*, 10 (2012), 2203–2214.

[55] Shammas, Sarah L, Travis, Alexandra J, and Clarke, Jane. Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cmyb to cbp kix. *The journal of physical chemistry B 117*, 42 (2013), 13346–13356.

[56] Huang, Yongqi, and Liu, Zhirong. Kinetic advantage of intrinsically disordered proteins in coupled folding–binding process: a critical assessment of the fly-casting mechanism. *Journal of molecular biology 393*, 5 (2009), 1143–1159.

[57] Shoemaker, Benjamin A, Portman, John J, and Wolynes, Peter G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proceedings of the National Academy of Sciences 97*, 16 (2000), 8868–8873.

[58] Zhou, Huan-Xiang, and Bates, Paul A. Modeling protein association mechanisms and kinetics. *Current opinion in structural biology 23*, 6 (2013), 887–893.

[59] Schor, Marieke, Mey, Antonia SJS, and MacPhee, Cait E. Analytical methods for structural ensembles and dynamics of intrinsically disordered proteins. *Biophysical reviews 8*, 4 (2016), 429–439.

[60] Dyson, H Jane, and Wright, Peter E. Unfolded proteins and protein folding studied by NMR. *Chemical reviews 104*, 8 (2004), 3607–3622.

[61] Palmer III, Arthur G. NMR characterization of the dynamics of biomacromolecules. *Chemical reviews 104*, 8 (2004), 3623–3640.

[62] Jensen, Malene Ringkjøbing, Ruigrok, Rob WH, and Blackledge, Martin. Describing intrinsically disordered proteins at atomic resolution by nmr. *Current opinion in structural biology 23*, 3 (2013), 426–435.

[63] Schwarzinger, Stephan, Kroon, Gerard JA, Foss, Ted R, Chung, John, Wright, Peter E, and Dyson, H Jane. Sequence-dependent correction of random coil nmr chemical shifts. *Journal of the American Chemical Society 123*, 13 (2001), 2970–2978.

[64] Tollinger, Martin, Skrynnikov, Nikolai R, Mulder, Frans AA, Forman-Kay, Julie D, and Kay, Lewis E. Slow dynamics in folded and unfolded states of an sh3 domain. *Journal of the American Chemical Society 123*, 46 (2001), 11341–11352.

[65] Sugita, Yuji, and Okamoto, Yuko. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters 314*, 1-2 (1999), 141–151.

[66] Liu, Pu, Kim, Byungchan, Friesner, Richard A, and Berne, BJ. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences 102*, 39 (2005), 13749–13754.

[67] Torrie, Glenn M, and Valleau, John P. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics 23*, 2 (1977), 187–199.

[68] Laio, Alessandro, and Parrinello, Michele. Escaping free-energy minima. *Proceedings of the National Academy of Sciences 99*, 20 (2002), 12562–12566.

[69] Swendsen, Robert H, and Wang, Jian-Sheng. Replica monte carlo simulation of spin-glasses. *Physical review letters 57*, 21 (1986), 2607.

[70] Park, Sanghyun, Khalili-Araghi, Fatemeh, Tajkhorshid, Emad, and Schulten, Klaus. Free energy calculation from steered molecular dynamics simulations using jarzynskis equality. *The Journal of chemical physics 119*, 6 (2003), 3559–3566.

[71] Camilloni, Carlo, and Vendruscolo, Michele. Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *Journal of the American Chemical Society 136*, 25 (2014), 8982–8991.

[72] Hall, Traci M Tanaka. Multiple modes of rna recognition by zinc finger proteins. *Current opinion in structural biology 15*, 3 (2005), 367–373.

[73] Liang, Jian, Song, Wenjun, Tromp, Gail, Kolattukudy, Pappachan E, and Fu, Mingui. Genome-wide survey and expression profiling of ccch-zinc finger family reveals a functional module in macrophage activation. *PLoS One 3*, 8 (2008), e2880.

[74] Wang, Dong, Guo, Yinghui, Wu, Changai, Yang, Guodong, Li, Yingying, and Zheng, Chengchao. Genome-wide analysis of ccch zinc finger family in arabidopsis and rice. *BMC genomics 9*, 1 (2008), 44.

[75] Fu, Mingui, and Blackshear, Perry J. Rna-binding proteins in immune regulation: a focus on ccch zinc finger proteins. *Nature Reviews Immunology 17*, 2 (2017), 130.

[76] Barreau, Carine, Paillard, Luc, and Osborne, H Beverley. AU-rich elements and associated factors: are there unifying principles? *Nucleic acids research 33*, 22 (2005), 7138–7150.

[77] Caput, D_, Beutler, B, Hartog, K, Thayer, R, Brown-Shimer, S, and Cerami, A. Identification of a common nucleotide sequence in the 3′-untranslated region of mrna molecules specifying inflammatory mediators. *Proceedings of the National Academy of Sciences 83*, 6 (1986), 1670–1674.

[78] Khabar, Khalid SA. Hallmarks of cancer and au-rich elements. *Wiley Interdisciplinary Reviews: RNA 8*, 1 (2017), e1368.

[79] Herschman, Harvey R. Primary response genes induced by growth factors and tumor promoters. *Annual review of biochemistry 60*, 1 (1991), 281–319.

[80] Schiavi, Susan C, Belasco, Joel G, and Greenberg, Michael E. Regulation of proto-oncogene mrna stability. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer 1114*, 2-3 (1992), 95–106.

[81] Espel, Enric, Garcia-Sanz, Jose A, Aubert, Vincent, Menoud, Véronique, Sperisen, Peter, Fernández, Neus, and Spertini, François. Transcriptional and translational control of tnf-α gene expression in human monocytes by major histocompatibility complex class ii ligands. *European journal of immunology 26*, 10 (1996), 2417–2424.

[82] Crawford, Eric K, Ensor, Jeffery E, Kalvakolanu, Indira, and Hasday, Jeffrey D. The role of 3′ poly (A) tail metabolism in tumor necrosis factor-α regulation. *Journal of Biological Chemistry 272*, 34 (1997), 21120–21127.

[83] Conne, Béatrice, Stutz, André, and Vassalli, Jean-Dominique. The 3′ untranslated region of messenger RNA: a molecular hotspotfor pathology? *Nature medicine 6*, 6 (2000), 637.

[84] Audic, Yann, and Hartley, Rebecca S. Post-transcriptional regulation in cancer. *Biology of the Cell 96*, 7 (2004), 479–498.

[85] Blackshear, PJ. Tristetraprolin and other ccch tandem zinc-finger proteins in the regulation of mrna turnover, 2002.

[86] Varnum, BC, Ma, QF, Chi, TH, Fletcher, B, and Herschman, HR. The tis11 primary response gene is a member of a gene family that encodes proteins with a highly conserved sequence containing an unusual cys-his repeat. *Molecular and cellular biology 11*, 3 (1991), 1754–1758.

[87] Chen, Ching-Yi, Gherzi, Roberto, Ong, Shao-En, Chan, Edward L, Raijmakers, Reinout, Pruijn, Ger JM, Stoecklin, Georg, Moroni, Christoph, Mann, Matthias, and Karin, Michael. Au binding proteins recruit the exosome to degrade are-containing mrnas. *Cell 107*, 4 (2001), 451–464.

[88] Carballo, Ester, Lai, Wi S, and Blackshear, Perry J. Feedback inhibition of macrophage tumor necrosis factor-α production by tristetraprolin. *Science 281*, 5379 (1998), 1001–1005.

[89] Lai, Wi S, Carballo, Ester, Thorn, Judith M, Kennington, Elizabeth A, and Blackshear, Perry J. Interactions of ccch zinc-finger proteins with mrna. 1. binding of tristetraprolin-related zinc finger proteins to au-rich elements and destabilization of mrna. *Journal of Biological Chemistry* (2000).

[90] Hudson, Brian P, Martinez-Yamout, Maria A, Dyson, H Jane, and Wright, Peter E. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nature Structural and Molecular Biology 11*, 3 (2004), 257.

[91] Worthington, Mark T, Amann, Barbara T, Nathans, Daniel, and Berg, Jeremy M. Metal binding properties and secondary structure of the zinc-binding domain of nup475. *Proceedings of the National Academy of Sciences 93*, 24 (1996), 13754–13759.

[92] Lai, Wi S, Kennington, Elizabeth A, and Blackshear, Perry J. Interactions of ccch zinc-finger proteins with mrna. non-binding tristetraprolin mutants exert an inhibitory effect on degradation of au-rich element containing mrnas. *Journal of Biological Chemistry* (2002).

[93] Lai, Wi S, Carballo, Ester, Strum, Julie R, Kennington, Elizabeth A, Phillips, Ruth S, and Blackshear, Perry J. Evidence that tristetraprolin binds to au-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mrna. *Molecular and cellular biology 19*, 6 (1999), 4311–4323.

[94] Blackshear, Perry J, Lai, Wi S, Kennington, Elizabeth A, Brewer, Gary, Wilson, Gerald M, Guan, Xiaoju, and Zhou, Pei. Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates. *Journal of Biological Chemistry 278*, 22 (2003), 19947–19955.

[95] Lai, Wi S, Kennington, Elizabeth A, and Blackshear, Perry J. Tristetraprolin and its family members can promote the cell-free deadenylation of au-rich element-containing mrnas by poly (a) ribonuclease. *Molecular and cellular biology 23*, 11 (2003), 3798–3812.

[96] Deveau, Laura M, and Massi, Francesca. Three residues make an evolutionary switch for folding and RNA-destabilizing activity in the TTP family of proteins. *ACS chemical biology 11*, 2 (2015), 435–443.

[97] Sanduja, Sandhya, Blanco, Fernando F, Young, Lisa E, Kaza, Vimala, and Dixon, Dan A. The role of tristetraprolin in cancer and inflammation. *Frontiers in bioscience: a journal and virtual library 17* (2012), 174.

[98] Stumpo, Deborah J, Broxmeyer, Hal E, Ward, Toni, Cooper, Scott, Hangoc, Giao, Chung, Yang Jo, Shelley, William C, Richfield, Eric K, Ray, Manas K, Yoder, Mervin C, et al. Targeted disruption of zfp36l2, encoding a ccch tandem zinc finger rna-binding protein, results in defective hematopoiesis. *Blood 114*, 12 (2009), 2401–2410.

[99] Iwanaga, Eisaku, Nanri, Tomoko, Mitsuya, Hiroaki, and Asou, Norio. Mutation in the rna binding protein TIS11D/ZFP36L2 is associated with the pathogenesis of acute leukemia. *International journal of oncology 38*, 1 (2011), 25–31.

[100] Deveau, Laura M. Characterizing the disorder in tristetraprolin and its contribution to post-transcriptional gene regulation: A dissertation.

[101] Seydoux, Geraldine, and Fire, Andrew. Soma-germline asymmetry in the distributions of embryonic rnas in caenorhabditis elegans. *Development 120*, 10 (1994), 2823–2834.

[102] Farley, Brian M, Pagano, John M, and Ryder, Sean P. RNA target specificity of the embryonic cell fate determinant POS-1. *Rna* (2008).

[103] Wallenfang, Matthew R, and Seydoux, Geraldine. Polarization of the anterior–posterior axis of c. elegans is a microtubule-directed process. *Nature 408*, 6808 (2000), 89.

[104] Griffin, Erik E. Cytoplasmic localization and asymmetric division in the early embryo of Caenorhabditis elegans. *Wiley Interdisciplinary Reviews: Developmental Biology 4*, 3 (2015), 267–282.

[105] Sulston, John E, Schierenberg, E, White, John G, and Thomson, JN. The embryonic cell lineage of the nematode caenorhabditis elegans. *Developmental biology 100*, 1 (1983), 64–119.

[106] Goldstein, Bob, and Macara, Ian G. The PAR proteins: fundamental players in animal cell polarization. *Developmental cell 13*, 5 (2007), 609–622.

[107] Kemphues, Kenneth. Parsing embryonic polarity. *Cell 101*, 4 (2000), 345–348.

[108] Schubert, Charlotte M, Lin, Rueyling, de Vries, Corry J, Plasterk, Ronald HA, and Priess, James R. MEX-5 and MEX-6 function to establish soma/germline asymmetry in early c. elegans embryos. *Molecular cell 5*, 4 (2000), 671–682.

[109] Wang, Jennifer T, and Seydoux, Geraldine. Germ cell specification. In *Germ Cell Development in C. elegans*. Springer, 2013, pp. 17–39.

[110] Pagano, John M, Farley, Brian M, McCoig, Lisa M, and Ryder, Sean P. Molecular basis of RNA recognition by the embryonic polarity determinant MEX-5. *Journal of Biological Chemistry* (2007).

[111] Griffin, Erik E, Odde, David J, and Seydoux, Geraldine. Regulation of the MEX-5 gradient by a spatially segregated kinase/phosphatase cycle. *Cell 146*, 6 (2011), 955–968.

[112] Cuenca, Adrian A, Schetter, Aaron, Aceto, Donato, Kemphues, Kenneth, and Seydoux, Geraldine. Polarization of the c. elegans zygote proceeds via distinct establishment and maintenance phases. *Development 130*, 7 (2003), 1255–1265.

[113] DeRenzo, Cynthia, Reese, Kimberly J, and Seydoux, Geraldine. Exclusion of germ plasm proteins from somatic lineages by cullin-dependent degradation. *Nature 424*, 6949 (2003), 685.

[114] Reese, Kimberly J, Dunn, Melanie A, Waddle, James A, and Seydoux, Geraldine. Asymmetric segregation of pie-1 in c. elegans is mediated by two complementary mechanisms that act through separate pie-1 protein domains. *Molecular cell 6*, 2 (2000), 445–455.

[115] Han, Bingjie, Antkowiak, Katianna R, Fan, Xintao, Rutigliano, Mallory, Ryder, Sean P, and Griffin, Erik E. Polo-like kinase couples cytoplasmic protein gradients in the C. elegans zygote. *Current Biology 28*, 1 (2018), 60–69.

[116] Trcek, Tatjana, and Lehmann, Ruth. Germ granules: All about the RNA after all. *eLife 6* (2017), e24106.

[117] Smith, Jarrett, Calidas, Deepika, Schmidt, Helen, Lu, Tu, Rasoloson, Dominique, and Seydoux, Geraldine. Spatial patterning of p granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3. *Elife 5* (2016), e21337.

[118] Saha, Shambaditya, Weber, Christoph A, Nousch, Marco, Adame-Arana, Omar, Hoege, Carsten, Hein, Marco Y, Osborne-Nishimura, Erin, Mahamid, Julia, Jahnel, Marcus, Jawerth, Louise, et al. Polar positioning of phase-separated liquid compartments in cells regulated by an mRNA competition mechanism. *Cell 166*, 6 (2016), 1572–1584.

[119] Pasinelli, Piera, and Brown, Robert H. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nature Reviews Neuroscience 7*, 9 (2006), 710.

[120] Gros-Louis, Francois, Gaspar, Claudia, and Rouleau, Guy A. Genetics of familial and sporadic amyotrophic lateral sclerosis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1762*, 11-12 (2006), 956–972.

[121] Leigh, PN, Whitwell, H, Garofalo, O, Buller, J, Swash, M, Martin, JE, Gallo, J-M, Weller, RO, and Anderton, BH. Ubiquitin-immunoreactive intraneuronal inclusions in amyotrophic lateral sclerosis: morphology, distribution, and specificity. *Brain 114*, 2 (1991), 775–788.

[122] Neumann, Manuela, Sampathu, Deepak M, Kwong, Linda K, Truax, Adam C, Micsenyi, Matthew C, Chou, Thomas T, Bruce, Jennifer, Schuck, Theresa, Grossman, Murray, Clark, Christopher M, McCluskey, Leo F, Miller, Bruce L, Masliah, Eliezer, Mackenzie, Ian R, Feldman, Howard, Feiden, Wolfgang, Kretzschmar, Hans A, Trojanowski, John Q, and Lee, Virginia M-Y. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science 314*, 5796 (2006), 130–133.

[123] Pesiridis, G Scott, Lee, Virginia M-Y, and Trojanowski, John Q. Mutations in tdp-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Human molecular genetics 18*, R2 (2009), R156–R162.

[124] Geser, Felix, Lee, Virginia M-Y, and Trojanowski, John Q. Amyotrophic lateral sclerosis and frontotemporal lobar degeneration: a spectrum of tdp-43 proteinopathies. *Neuropathology 30*, 2 (2010), 103–112.

[125] Ross, Christopher A, and Poirier, Michelle A. What is the role of protein aggregation in neurodegeneration? *Nature reviews Molecular cell biology 6*, 11 (2005), 891.

[126] Sunde, Margaret, and Blake, Colin CF. From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Quarterly reviews of biophysics 31*, 1 (1998), 1–39.

[127] Jucker, Mathias, and Walker, Lary C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature 501*, 7465 (2013), 45.

[128] Maji, Samir K, Perrin, Marilyn H, Sawaya, Michael R, Jessberger, Sebastian, Vadodaria, Krishna, Rissman, Robert A, Singru, Praful S, Nilsson, K Peter R, Simon, Rozalyn, Schubert, David, et al. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science 325*, 5938 (2009), 328–332.

[129] Blancas-Mejía, Luis M, and Ramirez-Alvarado, Marina. Systemic amyloidoses. *Annual review of biochemistry 82* (2013), 745–774.

[130] Eisenberg, David, and Jucker, Mathias. The amyloid state of proteins in human diseases. *Cell 148*, 6 (2012), 1188–1203.

[131] Mackenzie, Ian RA. The neuropathology of ftd associated with als. *Alzheimer Disease & Associated Disorders 21*, 4 (2007), S44–S49.

[132] Maekawa, Satomi, Leigh, P Nigel, King, Andrew, Jones, Edith, Steele, John C, Bodi, Istvan, Shaw, Christopher E, Hortobagyi, Tibor, and Al-Sarraj, Safa. Tdp-43 is consistently co-localized with ubiquitinated inclusions in sporadic and guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without sod1 mutations. *Neuropathology 29*, 6 (2009), 672–683.

[133] Tateishi, Takahisa, Hokonohara, Toshihiro, Yamasaki, Ryo, Miura, Shiro, Kikuchi, Hitoshi, Iwaki, Akiko, Tashiro, Hiroshi, Furuya, Hirokazu, Nagara, Yuko, Ohyagi, Yasumasa, et al. Multiple system degeneration with basophilic inclusions in japanese als patients with fus mutation. *Acta neuropathologica 119*, 3 (2010), 355–364.

[134] Scotter, Emma L, Chen, Han-Jou, and Shaw, Christopher E. Tdp-43 proteinopathy and als: insights into disease mechanisms and therapeutic targets. *Neurotherapeutics 12*, 2 (2015), 352–363.

[135] Wils, Hans, Kleinberger, Gernot, Janssens, Jonathan, Pereson, Sandra, Joris, Geert, Cuijt, Ivy, Smits, Veerle, Ceuterick-de Groote, Chantal, Van Broeckhoven, Christine, and Kumar-Singh, Samir. Tdp-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of als and frontotemporal lobar degeneration. *Proceedings of the National Academy of Sciences 107*, 8 (2010), 3858–3863.

[136] Li, Yan, Ray, Payal, Rao, Elizabeth J, Shi, Chen, Guo, Weirui, Chen, Xiaoping, Woodruff, Elvin A, Fushimi, Kazuo, and Wu, Jane Y. A drosophila model for tdp-43 proteinopathy. *Proceedings of the National Academy of Sciences 107*, 7 (2010), 3169–3174.

[137] Ash, Peter EA, Zhang, Yong-Jie, Roberts, Christine M, Saldi, Tassa, Hutter, Harald, Buratti, Emanuele, Petrucelli, Leonard, and Link, Christopher D. Neurotoxic effects of tdp-43 overexpression in c. elegans. *Human molecular genetics 19*, 16 (2010), 3206–3218.

[138] Wang, I-Fang, Guo, Bo-Shen, Liu, Yu-Chih, Wu, Cheng-Chun, Yang, Chun-Hung, Tsai, Kuen-Jer, and Shen, Che-Kun James. Autophagy activators rescue and alleviate pathogenesis of a mouse model with proteinopathies of the tar dna-binding protein 43. *Proceedings of the National Academy of Sciences 109*, 37 (2012), 15024–15029.

[139] Iguchi, Yohei, Katsuno, Masahisa, Niwa, Jun-ichi, Takagi, Shinnosuke, Ishigaki, Shinsuke, Ikenaka, Kensuke, Kawai, Kaori, Watanabe, Hirohisa, Yamanaka, Koji, Takahashi, Ryosuke, et al. Loss of tdp-43 causes age-dependent progressive motor neuron degeneration. *Brain 136*, 5 (2013), 1371–1382.

[140] Kabashi, Edor, Champagne, Nathalie, Brustein, Edna, and Drapeau, Pierre. In the swim of things: recent insights to neurogenetic disorders from zebrafish. *Trends in genetics 26*, 8 (2010), 373–381.

[141] Yang, Chunxing, Wang, Hongyan, Qiao, Tao, Yang, Bin, Aliaga, Leonardo, Qiu, Linghua, Tan, Weijia, Salameh, Johnny, McKenna-Yasek, Diane M, Smith, Thomas, et al. Partial loss of tdp-43 function causes phenotypes of amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences* (2014), 201322641.

[142] Diaper, Danielle C, Adachi, Yoshitsugu, Lazarou, Luke, Greenstein, Max, Simoes, Fabio A, Di Domenico, Angelique, Solomon, Daniel A, Lowe, Simon, Alsubaie, Rawan, Cheng, Daryl, et al. Drosophila tdp-43 dysfunction in glia and muscle cells cause cytological and behavioural phenotypes that characterize als and ftld. *Human molecular genetics 22*, 19 (2013), 3883–3893.

[143] Mackness, Brian C, Tran, Meme T, McClain, Shannan P, Matthews, C Robert, and Zitzewitz, Jill A. Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *Journal of Biological Chemistry 289*, 12 (2014), 8264–8276.

[144] Buratti, Emanuele, and Baralle, Francisco E. Characterization and functional implications of the rna binding properties of nuclear factor tdp-43, a novel splicing regulator of cftr exon 9. *Journal of Biological Chemistry* (2001).

[145] Tollervey, James R, Curk, Tomaž, Rogelj, Boris, Briese, Michael, Cereda, Matteo, Kayikci, Melis, König, Julian, Hortobágyi, Tibor, Nishimura, Agnes L, Župunski, Vera, et al. Characterizing the rna targets and position-dependent splicing regulation by tdp-43. *Nature neuroscience 14*, 4 (2011), 452.

[146] Bose, Jayarama Krishnan, Wang, I-Fan, Hung, Li, Tarn, Woan-Yuh, and Shen, C-K James. Tdp-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mrna splicing. *Journal of Biological Chemistry 283*, 43 (2008), 28852–28859.

[147] Mercado, Pablo Arrisi, Ayala, Youhna M, Romano, Maurizio, Buratti, Emanuele, and Baralle, Francisco E. Depletion of tdp 43 overrides the need for exonic and intronic splicing enhancers in the human apoa-ii gene. *Nucleic acids research 33*, 18 (2005), 6000–6010.

[148] Dreumont, Natacha, Bourgeois, Cyril F, Lejeune, Fabrice, Liu, Yilei, Ehrmann, Ingrid E, Elliott, David J, and Stévenin, James. Human rbmy regulates germline-specific splicing events by modulating the function of the serine/arginine-rich proteins 9g8 and tra2-β. *J Cell Sci 123*, 1 (2010), 40–50.

[149] Cohen, Todd J, Lee, Virginia MY, and Trojanowski, John Q. Tdp-43 functions and pathogenic mechanisms implicated in tdp-43 proteinopathies. *Trends in molecular medicine 17*, 11 (2011), 659–667.

[150] Elvira, George, Wasiak, Sylwia, Blandford, Vanessa, Tong, Xin-Kang, Serrano, Alexandre, Fan, Xiaotang, del Rayo Sánchez-Carbente, Maria, Servant, Florence, Bell, Alexander W, Boismenu, Daniel, et al. Characterization of an rna granule from developing brain. *Molecular & cellular proteomics 5*, 4 (2006), 635–651.

[151] Ou, SH, Wu, Foon, Harrich, David, García-Martínez, Leon F, and Gaynor, Richard B. Cloning and characterization of a novel cellular protein, tdp-43, that binds to human immunodeficiency virus type 1 tar dna sequence motifs. *Journal of virology 69*, 6 (1995), 3584–3596.

[152] Abhyankar, Mayuresh M, Urekar, Craig J, and Reddi, Prabhakara P. A novel cpg-free vertebrate insulator silences the testis-specific sp-10 gene in somatic tissues: role for tdp-43 in insulator function. *Journal of Biological Chemistry* (2007).

[153] Rutherford, Nicola J, Zhang, Yong-Jie, Baker, Matt, Gass, Jennifer M, Finch, NiCole A, Xu, Ya-Fei, Stewart, Heather, Kelley, Brendan J, Kuntz, Karen, Crook, Richard JP, et al. Novel mutations in tardbp (tdp-43) in patients with familial amyotrophic lateral sclerosis. *PLoS genetics 4*, 9 (2008), e1000193.

[154] Suzuki, Hiroaki, Lee, Kikyo, and Matsuoka, Masaaki. Tdp-43-induced death is associated with altered regulation of bim and bcl-xl and attenuated by caspase-mediated tdp-43 cleavage. *Journal of Biological Chemistry* (2011), jbc–M110.

[155] Huang, Chi-Chen, Bose, Jayarama Krishnan, Majumder, Pritha, Lee, Kuen-Haur, Huang, Jen-Tse Joseph, Huang, Jeffrey K, and Shen, Che-Kun James. Metabolism and mis-metabolism of the neuropathological signature protein tdp-43. *J Cell Sci* (2014), jcs–136150.

[156] Igaz, Lionel M, Kwong, Linda K, Lee, Edward B, Chen-Plotkin, Alice, Swanson, Eric, Unger, Travis, Malunda, Joe, Xu, Yan, Winton, Matthew J, Trojanowski, John Q, et al. Dysregulation of the als-associated gene tdp-43 leads to neuronal death and degeneration in mice. *The Journal of clinical investigation 121*, 2 (2011), 726–738.

[157] Zhang, Yong-Jie, Xu, Ya-Fei, Cook, Casey, Gendron, Tania F, Roettges, Paul, Link, Christopher D, Lin, Wen-Lang, Tong, Jimei, Castanedes-Casey, Monica, Ash, Peter, Gass, Jennifer, Rangachari, Vijayaraghavan, Buratti, Emanuele, Baralle, Francisco, Golde, Todd E., Dickson, Dennis W., and Petrucelli, Leonard. Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. *Proceedings of the National Academy of Sciences 106*, 18 (2009), 7607–7612.

[158] Johnson, Brian S, McCaffery, J Michael, Lindquist, Susan, and Gitler, Aaron D. A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proceedings of the National Academy of Sciences 105*, 17 (2008), 6439–6444.

[159] Wang, Yi-Ting, Kuo, Pan-Hsien, Chiang, Chien-Hao, Liang, Jhe-Ruei, Chen, Yun-Ru, Wang, Shuying, Shen, James CK, and Yuan, Hanna S. The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *Journal of Biological Chemistry 288*, 13 (2013), 9049–9057.

[160] Voigt, Aaron, Herholz, David, Fiesel, Fabienne C, Kaur, Kavita, Müller, Daniel, Karsten, Peter, Weber, Stephanie S, Kahle, Philipp J, Marquardt, Till, and Schulz, Jörg B. Tdp-43-mediated neuron loss in vivo requires rna-binding activity. *PloS one 5*, 8 (2010), e12247.

[161] Dormann, Dorothee, Capell, Anja, Carlson, Aaron M, Shankaran, Sunita S, Rodde, Ramona, Neumann, Manuela, Kremmer, Elisabeth, Matsuwaki, Takashi, Yamanouchi, Keitaro, Nishihara, Masugi, et al. Proteolytic processing of tar dna binding protein-43 by caspases produces c-terminal fragments with disease defining properties independent of progranulin. *Journal of neurochemistry 110*, 3 (2009), 1082–1094.

[162] Arnold, Eveline S, Ling, Shuo-Chien, Huelga, Stephanie C, Lagier-Tourenne, Clotilde, Polymenidou, Magdalini, Ditsworth, Dara, Kordasiewicz, Holly B, McAlonis-Downes, Melissa, Platoshyn, Oleksandr, Parone, Philippe A, et al. Als-linked tdp-43 mutations produce aberrant rna splicing and adult-onset motor neuron disease without aggregation or loss of nuclear tdp-43. *Proceedings of the National Academy of Sciences 110*, 8 (2013), E736–E745.

[163] Blackshear, P J. Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem. Soc. Trans. 30*, Pt 6 (Nov. 2002), 945–952.

[164] Carrick, Danielle M, Lai, Wi S, and Blackshear, Perry J. The tandem CCCH zinc finger protein tristetraprolin and its relevance to cytokine mRNA turnover and arthritis. *Arthritis Res. Ther. 6*, 6 (2004), 248–264.

[165] Carballo, Ester, Lai, Wi S, and Blackshear, Perry J. Feedback Inhibition of Macrophage Tumor Necrosis Factor-α Production by Tristetraprolin. *Science 281*, 5379 (Aug. 1998), 1001–1005.

[166] Ogilvie, Rachel L, Abelson, Michelle, Hau, Heidi H, Vlasova, Irina, Bohjanen, Paul R, and Blackshear, Perry J. Tristetraprolin Down-Regulates IL-2 Gene Expression through AU-Rich Element-Mediated mRNA Decay. *J. Immunol. 174*, 2 (Jan. 2005), 953–961.

[167] Lai, Wi S, Carballo, Ester, Strum, Julie R, Kennington, Elizabeth A, Phillips, Ruth S, and Blackshear, Perry J. Evidence that Tristetraprolin Binds to AU-Rich Elements and Promotes the Deadenylation and Destabilization of Tumor Necrosis Factor Alpha mRNA. *Molecular and Cellular Biology 19*, 6 (June 1999), 4311–4323.

[168] Chen, C Y, Gherzi, R, Ong, S E, Chan, E L, Raijmakers, R, Pruijn, G J, Stoecklin, G, Moroni, C, Mann, M, and Karin, M. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell 107*, 4 (Nov. 2001), 451–464.

[169] Lai, Wi S, Kennington, Elizabeth A, and Blackshear, Perry J. Interactions of CCCH zinc finger proteins with mRNA: non-binding tristetraprolin mutants exert an inhibitory effect on degradation of AU-rich element-containing mRNAs. *J. Biol. Chem. 277*, 11 (Mar. 2002), 9606–9613.

[170] Lai, W S, Carballo, E, Thorn, J M, Kennington, E A, and Blackshear, P J. Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. *J. Biol. Chem. 275*, 23 (June 2000), 17827–17837.

[171] Hudson, Brian P, Martinez-Yamout, Maria A, Dyson, H Jane, and Wright, Peter E. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat Struct Mol Biol 11*, 3 (Mar. 2004), 257–264.

[172] Morgan, Brittany R, and Massi, Francesca. A computational study of RNA binding and specificity in the tandem zinc finger domain of TIS11d. *Protein Sci. 19*, 6 (June 2010), 1222–1234.

[173] Morgan, Brittany R, Deveau, Laura M, and Massi, Francesca. Probing the structural and dynamical effects of the charged residues of the TZF domain of TIS11d. *Biophys. J. 108*, 6 (Mar. 2015), 1503–1515.

[174] Camilloni, Carlo, De Simone, Alfonso, Vranken, Wim F, and Vendruscolo, Michele. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry 51*, 11 (Mar. 2012), 2224–2231.

[175] Cornilescu, G, Delaglio, F, and Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR 13*, 3 (Mar. 1999), 289–302.

[176] Deveau, Laura M, and Massi, Francesca. Three residues make an evolutionary switch for folding and RNA-destabilizing activity in the TTP family of proteins. *ACS Chem. Biol.* (Nov. 2015).

[177] Brewer, Brandy Y, Ballin, Jeff D, Fialcowitz-White, Elizabeth J, Blackshear, Perry J, and Wilson, Gerald M. Substrate dependence of conformational changes in the RNA-binding domain of tristetraprolin assessed by fluorescence spectroscopy of tryptophan mutants. *Biochemistry 45*, 46 (Nov. 2006), 13807–13817.

[178] Blackshear, Perry J, Lai, Wi S, Kennington, Elizabeth A, Brewer, Gary, Wilson, Gerald M, Guan, Xiaoju, and Zhou, Pei. Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates. *J. Biol. Chem. 278*, 22 (May 2003), 19947–19955.

[179] Ballesteros, J A, Deupi, X, Olivella, M, and Haaksma, EEJ. Serine and threonine residues bend α-helices in the χ 1= g- conformation. *Biophys. J. 79*, 3 (2000), 2754–2760.

[180] Finn, Robert D, Bateman, Alex, Clements, Jody, Coggill, Penelope, Eberhardt, Ruth Y, Eddy, Sean R, Heger, Andreas, Hetherington, Kirstie, Holm, Liisa, Mistry, Jaina, Sonnhammer, Erik L L, Tate, John, and Punta, Marco. Pfam: the protein families database. *Nucleic Acids Research 42*, Database issue (Jan. 2014), D222–30.

[181] Sakharov, D V, and Lim, C. Zn protein simulations including charge transfer and local polarization effects. *J. Am. Chem. Soc. 127* (2004), 4921–4929.

[182] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A 38* (1988), 3098–3100.

[183] Lee, Chengteh, Yang, Weitao, and Parr, Robert G. Development of the Calle-Salvetti correlation energy into a functional of the electron density. *Phys. Rev. B 37* (1988), 785–789.

[184] Becke, Axel D. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys. 98* (1993), 5648–5652.

[185] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

[186] MacKerell, A D, Bashford, D, Bellott, M, Dunbrack, R L, Evanseck, J D, Field, M J, Fischer, S, Gao, J, Guo, H, Ha, S, Joseph-McCarthy, D, Kuchnir, L, Kuczera, K, Lau, F T K, Mattos, C, Michnick, S, Ngo, T, Nguyen, D T, Prodhom, B, Reiher, W E, Roux, B, Schlenkrich, M, Smith, J C, Stote, R, Straub, J, Watanabe, M, Wiorkiewicz-Kuczera, J, Yin, D, and Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B J Phys Chem B 102*, 18 (1998), 3586–3616.

[187] Foster, J. P., and Weinhold, F. Natural hybrid orbitals. *J. Am. Chem. Soc. 102* (1980), 7211–7218.

[188] Reed, A. E., and Weinhold, F. Natural bond orbital analysis of near-hartree-fock water dimer. *J. Chem. Phys. 78* (1983), 4066–4073.

[189] Reed, A. E., Weinstock, R. B., and Weinhold, F. Natural-population analysis. *J. Chem. Phys. 83* (1985), 735–746.

[190] Reed, A. E., and Weinhold, F. Natural localized molecular orbitals. *J. Chem. Phys. 83* (1985), 1736–1740.

[191] Bordoli, Lorenza, Kiefer, Florian, Arnold, Konstantin, Benkert, Pascal, Battey, James, and Schwede, Torsten. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc 4*, 1 (2009), 1–13.

[192] Arnold, Konstantin, Bordoli, Lorenza, Kopp, Jürgen, and Schwede, Torsten. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics 22*, 2 (Jan. 2006), 195–201.

[193] Schwede, Torsten, Kopp, Jürgen, Guex, Nicolas, and Peitsch, Manuel C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res. 31*, 13 (July 2003), 3381–3385.

[194] Humphrey, William, Dalke, Andrew, and Schulten, Klaus. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics 14* (1996), 33–38.

[195] Phillips, James C, Braun, Rosemary, Wang, Wei, Gumbart, James, Tajkhorshid, Emad, Villa, Elizabeth, Chipot, Christophe, Skeel, Robert D, Kalé, Laxmikant, and Schulten, Klaus. Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 16 (Dec. 2005), 1781–1802.

[196] Ryckaert, J P, Ciccotti, G, and Berendsen, H J C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23* (1997), 327–341.

[197] Darden, Tom, York, Darrin M, and Pedersen, Lee G. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys. 96*, 12 (June 1993), 10089–10092.

[198] Essmann, Ulrich, Perera, Lalith, Berkowitz, Max L, Darden, Tom, Lee, Hsing, and Pedersen, Lee G. A smooth particle mesh Ewald method. *J. Chem. Phys. 103*, 19 (Nov. 1995), 8577–8593.

[199] Frishman, D., and Argos, P. Knowledge-based secondary structure assignment. *Proteins: structure, function and genetics 23* (1995), 566–579.

[200] Stone, John. *An Efficient Library for Parallel Ray Tracing and Animation.* Master's thesis, Computer Science Department, University of Missouri-Rolla, April 1998.

[201] Goddard, T. D., and Kneller, D. G. SPARKY 3. University of California, San Francisco.

[202] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. NMRPipe - a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR 6* (1995), 277–293.

[203] Hwang, Sue-Yun, and Rose, Lesilee S. Control of asymmetric cell division in early C. elegans embryogenesis: teaming-up translational repression and protein degradation. *BMB reports 43*, 2 (2010), 69.

[204] Kumano, Gaku. Polarizing animal cells via mRNA localization in oogenesis and early development. *Development, growth & differentiation 54*, 1 (2012), 1–18.

[205] St Johnston, Daniel, and Ahringer, Julie. Cell polarity in eggs and epithelia: parallels and diversity. *Cell 141*, 5 (2010), 757–774.

[206] Rose, Lesilee S, and Kemphues, Kenneth J. Early patterning of the C. elegans embryo. *Annual review of genetics 32*, 1 (1998), 521–545.

[207] Tenlen, Jennifer R, Molk, Jeffrey N, London, Nitobe, Page, Barbara D, and Priess, James R. MEX-5 asymmetry in one-cell c. elegans embryos requires PAR-4-and PAR-1-dependent phosphorylation. *Development 135*, 22 (2008), 3665–3675.

[208] Brown, Guy C, and Kholodenko, Boris N. Spatial gradients of cellular phosphoproteins. *FEBS letters 457*, 3 (1999), 452–454.

[209] Lipkow, Karen, and Odde, David J. Model for protein concentration gradients in the cytoplasm. *Cellular and molecular bioengineering 1*, 1 (2008), 84–92.

[210] Budirahardja, Yemima, and Gönczy, Pierre. PLK-1 asymmetry contributes to asynchronous cell division of C. elegans embryos. *Development 135*, 7 (2008), 1303–1313.

[211] Mello, Craig C, Schubert, Charlotte, Draper, Bruce, Zhang, Wei, Lobel, Robert, and Priess, James R. The PIE-1 protein and germline specification in C. elegans embryos. *Nature 382*, 6593 (1996), 710.

[212] Rivers, David M, Moreno, Sergio, Abraham, Mary, and Ahringer, Julie. PAR proteins direct asymmetry of the cell cycle regulators Polo-like kinase and Cdc25. *The Journal of cell biology 180*, 5 (2008), 877–885.

[213] Gallo, Christopher M, Munro, Edwin, Rasoloson, Dominique, Merritt, Christopher, and Seydoux, Geraldine. Processing bodies and germ granules are distinct RNA granules that interact in C. elegans embryos. *Developmental biology 323*, 1 (2008), 76–87.

[214] Varnum, BC, Lim, RW, Kujubu, DA, Luner, SJ, Kaufman, SE, Greenberger, JS, Gasson, JC, and Herschman, HR. Granulocyte-macrophage colony-stimulating factor and tetradecanoyl phorbol acetate induce a distinct, restricted subset of primary-response TIS genes in both proliferating and terminally differentiated myeloid cells. *Molecular and cellular biology 9*, 8 (1989), 3580–3583.

[215] DuBois, Raymond N, McLane, Michael W, Ryder, Kevin, Lau, Lester F, and Nathans, Daniel. A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence. *Journal of Biological Chemistry 265*, 31 (1990), 19185–19191.

[216] Lai, Wi S, Stumpo, DJ, and Blackshear, PJ. Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein. *Journal of Biological Chemistry 265*, 27 (1990), 16556–16563.

[217] Kaymak, Ebru, and Ryder, Sean P. RNA recognition by the C. elegans oocyte maturation determinant OMA-1. *Journal of Biological Chemistry* (2013), jbc–M113.

[218] Tavella, Davide, Deveau, Laura M, Whitfield, Troy W, and Massi, Francesca. Structural basis of the disorder in the tandem zinc finger domain of the rna-binding protein tristetraprolin. *Journal of chemical theory and computation 12*, 10 (2016), 4717–4725.

[219] Bhattacharya, Aneerban, Tejero, Roberto, and Montelione, Gaetano T. Evaluating protein structures determined by structural genomics consortia. *Proteins: Structure, Function, and Bioinformatics 66*, 4 (2007), 778–795.

[220] Dembinski, Holly, Wismer, Kevin, Balasubramaniam, Deepa, Gonzalez, Hector A, Alverdi, Vera, Iakoucheva, Lilia M, and Komives, Elizabeth A. Predicted disorder-to-order transition mutations in IκBα disrupt function. *Physical Chemistry Chemical Physics 16*, 14 (2014), 6480–6485.

[221] Güntert, Peter, and Buchner, Lena. Combined automated noe assignment and structure calculation with cyana. *Journal of biomolecular NMR 62*, 4 (2015), 453–471.

[222] Shen, Yang, Delaglio, Frank, Cornilescu, Gabriel, and Bax, Ad. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR 44*, 4 (2009), 213–223.

[223] Brünger, Axel T, Adams, Paul D, Clore, G Marius, DeLano, Warren L, Gros, Piet, Grosse-Kunstleve, Ralf W, Jiang, Jian-Sheng, Kuszewski, John, Nilges, Michael, Pannu, Navraj S, et al. Crystallography & nmr system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D 54*, 5 (1998), 905–921.

[224] Brunger, Axel T. Version 1.2 of the crystallography and nmr system. *Nature protocols 2*, 11 (2007), 2728.

[225] Lee, Edward B, Lee, Virginia M-Y, and Trojanowski, John Q. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nature Reviews Neuroscience 13*, 1 (2012), 38.

[226] Buratti, Emanuele, Dörk, Thilo, Zuccato, Elisabetta, Pagani, Franco, Romano, Maurizio, and Baralle, Francisco E. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *The EMBO journal 20*, 7 (2001), 1774–1784.

[227] Buratti, Emanuele, and Baralle, Francisco E. Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. *Frontiers in Bioscience 13*, 867-7 (2008), 8.

[228] Arai, Tetsuaki, Hasegawa, Masato, Akiyama, Haruhiko, Ikeda, Kenji, Nonaka, Takashi, Mori, Hiroshi, Mann, David, Tsuchiya, Kuniaki, Yoshida, Mari, Hashizume, Yoshio, and Oda, Tatsuro. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochemical and biophysical research communications 351*, 3 (2006), 602–611.

[229] Abel, Olubunmi, Shatunov, Aleksey, Jones, Ashley R, Andersen, Peter M, Powell, John F, and Al-Chalabi, Ammar. Development of a smartphone app for a genetics website: the amyotrophic lateral sclerosis online genetics database (alsod). *JMIR mHealth and uHealth 1*, 2 (2013).

[230] Johnson, Brian S, Snead, David, Lee, Jonathan J, McCaffery, J Michael, Shorter, James, and Gitler, Aaron D. TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *Journal of Biological Chemistry 284*, 30 (2009), 20329–20339.

[231] Igaz, Lionel M, Kwong, Linda K, Xu, Yan, Truax, Adam C, Uryu, Kunihiro, Neumann, Manuela, Clark, Christopher M, Elman, Lauren B, Miller, Bruce L, Grossman, Murray, McCluskey, Leo F, Trojanowski, John Q, and Lee, Virginia M-Y. Enrichment of C-terminal fragments in TAR DNA-binding protein-43 cytoplasmic inclusions in brain but not in spinal cord of frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *The American journal of pathology 173*, 1 (2008), 182–194.

[232] Zhang, Yong-Jie, Xu, Ya-Fei, Dickey, Chad A, Buratti, Emanuele, Baralle, Francisco, Bailey, Rachel, Pickering-Brown, Stuart, Dickson, Dennis, and Petrucelli, Leonard. Progranulin mediates caspase-dependent cleavage of TAR DNA binding protein-43. *J. Neurosci. 27*, 39 (Sept. 2007), 10530–10534.

[233] Igaz, Lionel M, Kwong, Linda K, Chen-Plotkin, Alice, Winton, Matthew J, Unger, Travis L, Xu, Yan, Neumann, Manuela, Trojanowski, John Q, and Lee, Virginia M-Y. Expression of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies. *Journal of Biological Chemistry 284*, 13 (2009), 8516–8524.

[234] Yang, Chunxing, Tan, Weijia, Whittle, Catheryne, Qiu, Linghua, Cao, Lucheng, Akbarian, Schahram, and Xu, Zuoshang. The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism. *PloS one 5*, 12 (2010), e15878.

[235] Kuo, Pan-Hsien, Doudeva, Lyudmila G, Wang, Yi-Ting, Shen, Che-Kun James, and Yuan, Hanna S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic acids research 37*, 6 (2009), 1799–1808.

[236] Kathuria, Sagar V, Chan, Yvonne H, Nobrega, R Paul, Özen, Ayşegül, and Matthews, C Robert. Clusters of isoleucine, leucine, and valine side chains define cores of stability in high-energy states of globular proteins: Sequence determinants of structure and stability. *Protein Science 25*, 3 (2016), 662–675.

[237] Li, Quan, Yokoshi, Moe, Okada, Hitomi, and Kawahara, Yukio. The cleavage pattern of TDP-43 determines its rate of clearance and cytotoxicity. *Nature communications 6* (2015), 6183.

[238] Neudecker, Philipp, Robustelli, Paul, Cavalli, Andrea, Walsh, Patrick, Lundström, Patrik, Zarrine-Afsar, Arash, Sharpe, Simon, Vendruscolo, Michele, and Kay, Lewis E. Structure of an intermediate state in protein folding and aggregation. *Science 336*, 6079 (2012), 362–366.

[239] Morgan, Brittany R, Zitzewitz, Jill A, and Massi, Francesca. Structural rearrangement upon fragmentation of the stability core of the ALS-linked protein TDP-43. *Biophysical journal 113*, 3 (2017), 540–549.

[240] Le Marchand, T, de Rosa, M, Salvi, N, Sala, BM, Andreas, LB, Barbet-Massin, E, Sormanni, P, Barbiroli, A, Porcari, R, Sousa, Mota C, de Sanctis, D, Bolognesi, M, Emsley, L, Bellotti, V, Blackledge, M, Camilloni, C, Pintacuda, G, and Ricagno, S. Conformational dynamics in crystals reveal the molecular bases for D76N beta-2 microglobulin aggregation propensity. *Nature communications 9*, 1 (2018), 1658.

[241] Best, Robert B, Zhu, Xiao, Shim, Jihyun, Lopes, Pedro EM, Mittal, Jeetain, Feig, Michael, and MacKerell Jr, Alexander D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of chemical theory and computation 8*, 9 (2012), 3257–3273.

[242] Best, Robert B, Mittal, Jeetain, Feig, Michael, and MacKerell Jr, Alexander D. Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of α-helix and β-hairpin formation. *Biophysical journal 103*, 5 (2012), 1045–1051.

[243] Lindorff-Larsen, Kresten, Maragakis, Paul, Piana, Stefano, Eastwood, Michael P, Dror, Ron O, and Shaw, David E. Systematic validation of protein force fields against experimental data. *PloS one 7*, 2 (2012), e32131.

[244] Li, Da-Wei, and Brüschweiler, Rafael. Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. *Journal of chemical theory and computation 7*, 6 (2011), 1773–1782.

[245] Freddolino, Peter L, Harrison, Christopher B, Liu, Yanxin, and Schulten, Klaus. Challenges in protein-folding simulations. *Nature physics 6*, 10 (2010), 751.

[246] Camilloni, Carlo, Cavalli, Andrea, and Vendruscolo, Michele. Replica-averaged metadynamics. *Journal of chemical theory and computation 9*, 12 (2013), 5610–5617.

[247] Pitera, Jed W, and Chodera, John D. On the use of experimental observations to bias simulated ensembles. *Journal of Chemical Theory and Computation 8*, 10 (2012), 3445–3451.

[248] Cavalli, Andrea, Camilloni, Carlo, and Vendruscolo, Michele. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *The Journal of Chemical Physics 138*, 9 (2013), 03B603.

[249] Roux, Benoît, and Weare, Jonathan. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of Chemical Physics 138*, 8 (2013), 02B616.

[250] Piana, Stefano, and Laio, Alessandro. A bias-exchange approach to protein folding. *The journal of physical chemistry B 111*, 17 (2007), 4553–4559.

[251] Camilloni, Carlo, Robustelli, Paul, Simone, Alfonso De, Cavalli, Andrea, and Vendruscolo, Michele. Characterization of the conformational equilibrium between the two major substates of rnase a using NMR chemical shifts. *Journal of the American Chemical Society 134*, 9 (2012), 3968–3971.

[252] Mackness, Brian C, Morgan, Brittany R, Deveau, Laura M, Kathuria, Sagar V, Massi, Francesca, and Zitzewitz, Jill A. A hydrophobic core stabilizes the residual structure in the RRM2 intermediate state of the ALS-linked protein TDP-43. *under review*.

[253] Pace, C N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Meth. Enzymol. 131* (1986), 266–280.

[254] Bennion, Brian J, and Daggett, Valerie. The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences 100*, 9 (2003), 5142–5147.

[255] Mielke, Steven P, and Krishnan, VV. Characterization of protein secondary structure from NMR chemical shifts. *Progress in nuclear magnetic resonance spectroscopy 54*, 3-4 (2009), 141.

[256] Abrams, Cameron, and Bussi, Giovanni. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy 16*, 1 (2013), 163–199.

[257] Maragliano, Luca, and Vanden-Eijnden, Eric. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical physics letters 426*, 1 (2006), 168–175.

[258] Laio, Alessandro, and Gervasio, Francesco L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics 71*, 12 (2008), 126601.

[259] Zambrano, Rafael, Jamroz, Michal, Szczasiuk, Agata, Pujols, Jordi, Kmiecik, Sebastian, and Ventura, Salvador. AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic acids research 43*, W1 (2015), W306–W313.

[260] Guenther, Elizabeth L, Ge, Peng, Trinh, Hamilton, Sawaya, Michael R, Cascio, Duilio, Boyer, David R, Gonen, Tamir, Zhou, Z Hong, and Eisenberg, David S. Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nature structural & molecular biology 25*, 4 (2018), 311.

[261] Arai, Tetsuaki, Hasegawa, Masato, Nonoka, Takashi, Kametani, Fuyuki, Yamashita, Makiko, Hosokawa, Masato, Niizato, Kazuhiro, Tsuchiya, Kuniaki, Kobayashi, Zen, Ikeda, Kenji, Yoshida, Mari, Onaya, Mitsumoto, Fujishiro, Hiroshige, and Akiyama, Haruhiko. Phosphorylated and cleaved TDP-43 in ALS, FTLD and other neurodegenerative disorders and in cellular models of TDP-43 proteinopathy. *Neuropathology 30*, 2 (2010), 170–181.

[262] Winton, Matthew J, Igaz, Lionel M, Wong, Margaret M, Kwong, Linda K, Trojanowski, John Q, and Lee, Virginia M-Y. Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. *Journal of Biological Chemistry 283*, 19 (2008), 13302–13309.

[263] Cohen, Todd J, Hwang, Andrew W, Unger, Travis, Trojanowski, John Q, and Lee, Virginia MY. Redox signalling directly regulates TDP-43 via cysteine oxidation and disulphide cross-linking. *The EMBO journal 31*, 5 (2012), 1241–1252.

[264] Pesiridis, G Scott, Tripathy, Kalyan, Tanik, Selçuk, Trojanowski, John Q, and Lee, Virginia M-Y. A "two-hit" hypothesis for inclusion formation by carboxyl-terminal fragments of TDP-43 protein linked to RNA depletion and impaired microtubule-dependent transport. *Journal of Biological Chemistry 286*, 21 (2011), 18845–18855.

[265] Furukawa, Yoshiaki, Kaneko, Kumi, Watanabe, Shoji, Yamanaka, Koji, and Nukina, Nobuyuki. A seeding reaction recapitulates intracellular formation of sarkosyl-insoluble transactivation response element (TAR) dna-binding protein-43 inclusions. *Journal of Biological Chemistry 286*, 21 (2011), 18664–18672.

[266] Kabsch, Wolfgang, and Sander, Christian. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 12 (1983), 2577–2637.

[267] Pall, Szilárd, Abraham, Mark James, Kutzner, Carsten, Hess, Berk, and Lindahl, Erik. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In *International Conference on Exascale Applications and Software* (2014), Springer, pp. 3–27.

[268] Pronk, Sander, Páll, Szilárd, Schulz, Roland, Larsson, Per, Bjelkmar, Pär, Apostolov, Rossen, Shirts, Michael R, Smith, Jeremy C, Kasson, Peter M, Van Der Spoel, David, Hess, Berk, and Lindahl, Erik. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 7 (2013), 845–854.

[269] Hess, Berk, Kutzner, Carsten, Van Der Spoel, David, and Lindahl, Erik. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation 4*, 3 (2008), 435–447.

[270] Berendsen, Herman JC, van der Spoel, David, and van Drunen, Rudi. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications 91*, 1-3 (1995), 43–56.

[271] Best, Robert B, and Mittal, Jeetain. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *The Journal of Physical Chemistry B 114*, 46 (2010), 14916–14923.

[272] Abascal, Jose LF, and Vega, Carlos. A general purpose model for the condensed phases of water: TIP4P/2005. *The Journal of chemical physics 123*, 23 (2005), 234505.

[273] Hess, Berk, Bekker, Henk, Berendsen, Herman JC, and Fraaije, Johannes GEM. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry 18*, 12 (1997), 1463–1472.

[274] Darden, Tom, York, Darrin, and Pedersen, Lee. Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. *The Journal of chemical physics 98*, 12 (1993), 10089–10092.

[275] Evans, Denis J, and Holian, Brad Lee. The Nosé–Hoover thermostat. *The Journal of chemical physics 83*, 8 (1985), 4069–4074.

[276] Tribello, Gareth A, Bonomi, Massimiliano, Branduardi, Davide, Camilloni, Carlo, and Bussi, Giovanni. PLUMED 2: New feathers for an old bird. *Computer Physics Communications 185*, 2 (2014), 604–613.

[277] Camilloni, Carlo, Cavalli, Andrea, and Vendruscolo, Michele. Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins. *The Journal of Physical Chemistry B 117*, 6 (2013), 1838–1843.

[278] Robustelli, Paul, Kohlhoff, Kai, Cavalli, Andrea, and Vendruscolo, Michele. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure 18*, 8 (2010), 923–933.

[279] Kohlhoff, Kai J, Robustelli, Paul, Cavalli, Andrea, Salvatella, Xavier, and Vendruscolo, Michele. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *Journal of the American Chemical Society 131*, 39 (2009), 13894–13895.

[280] Pietrucci, Fabio, and Laio, Alessandro. A collective variable for the efficient exploration of protein beta-sheet structures: Application to SH3 and GB1. *Journal of Chemical Theory and Computation 5*, 9 (2009), 2197–2201.

[281] Biarnés, Xevi, Pietrucci, Fabio, Marinelli, Fabrizio, and Laio, Alessandro. METAGUI. a VMD interface for analyzing metadynamics and molecular dynamics simulations. *Computer Physics Communications 183*, 1 (2012), 203–211.

[282] Marinelli, Fabrizio, Pietrucci, Fabio, Laio, Alessandro, and Piana, Stefano. A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS computational biology 5*, 8 (2009), e1000452.

[283] Kumar, Shankar, Rosenberg, John M, Bouzida, Djamal, Swendsen, Robert H, and Kollman, Peter A. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational chemistry 13*, 8 (1992), 1011–1021.

[284] Rodriguez, Alex, and Laio, Alessandro. Clustering by fast search and find of density peaks. *Science 344*, 6191 (2014), 1492–1496.

[285] Van Der Lee, Robin, Buljan, Marija, Lang, Benjamin, Weatheritt, Robert J, Daughdrill, Gary W, Dunker, A Keith, Fuxreiter, Monika, Gough, Julian, Gsponer, Joerg, Jones, David T, et al. Classification of intrinsically disordered regions and proteins. *Chemical reviews 114*, 13 (2014), 6589–6631.

[286] Mackness, Brian C. The identification and targeting of partially-folded conformations on the folding free-energy landscapes of ALS-linked proteins for therapeutic intervention: A dissertation.

[287] Korzhnev, Dmitry M, Salvatella, Xavier, Vendruscolo, Michele, Di Nardo, Ariel A, Davidson, Alan R, Dobson, Christopher M, and Kay, Lewis E. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature 430*, 6999 (2004), 586.

[288] Korzhnev, Dmitry M, and Kay, Lewis E. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Accounts of Chemical Research 41*, 3 (2008), 442–451.

[289] Azzarito, Valeria, Long, Kérya, Murphy, Natasha S, and Wilson, Andrew J. Inhibition of α-helix-mediated protein–protein interactions using designed molecules. *Nature chemistry 5*, 3 (2013), 161.

[290] Zhang, Xiao-Hui, Tee, Louis Y, Wang, Xiao-Gang, Huang, Qun-Shan, and Yang, Shi-Hua. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy-Nucleic Acids 4* (2015).

[291] Sönnichsen, B, Koski, LB, Walsh, A, Marschall, P, Neumann, B, Brehm, M, Alleaume, A-M, Artelt, J, Bettencourt, P, Cassin, E, et al. Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature 434*, 7032 (2005), 462.